| ERIC Document Resume | See Instruction Sheet | DATE OF RESUME | DATE RECEIVED |
|---|---|---|---|

A. PUBLICATION DATA

1. ERIC NUMBER: ED001232
2. ORIGINATOR CONTROL NO(S).: N.A.
3. IDENTIFICATION NO.: CR 1790A

DO NOT WRITE IN SPACE BELOW

MICROFILM COPY AVAILABLE YES ☒ NO ☐

COPYRIGHT GRANTED YES ☐ NO ☒

COPYRIGHT RELEASE FILED YES ☐ NO ☒

TYPE OF REPORT
PROPOSAL ☐
IN-PROGRESS ☐
FINAL ☒
FOLLOW-UP ☐

OTHER ERIC REFERENCES:

ERIC SATELLITE CODE

4. AUTHOR(S): Anderson, Richard C.
5. TITLE OF DOCUMENT ABSTRACTED: Shaping Logical Behavior in Six and Seven-Year Olds
6. PUBLICATION TITLE: N.A.
7. EDITOR(S): N.A.
8. PLACE OF ORIGIN
University of Illinois
ORG. Urbana, Illinois
ST.
CITY, STATE
TELE. EXT.
INST/LOC CODE
9. PUBLISHER OR SOURCE Cooperative Research Program, Office of Education, Department of Health, Education, and Welfare, Wash. D.C.

| 10. VOL. | 11. PAGES | 12. MO./YR. | 13. TOTAL PP. |
|---|---|---|---|
| N.A. | N.A. | 1964 | 168 |

B. INDEXING DATA

14. SUBJECT AREA
15. TYPE OF ENDEAVOR report

16. INDEXING TERMS (6-12)

| | | | | |
|---|---|---|---|---|
| | | logical thinking | problem solving | |
| child development | instruction techniques | | programed instruction | |
| elementary school | learning theory | measurement | | |
| first grade | logical behavior | method of difference | | |

17. ABSTRACT (MAX. 250 WORDS)

A technology is needed to bridge the gap between educational theory and the practice of teaching. The chief goal was to develop a procedure for teaching children the Method of Difference, the problem-solving strategy in which each factor is varied in succession while all other factors are held constant. The components, scope, and limitations comprise a prescription that can furnish guidelines in the construction of a teaching program. The developed program consisted of seven large units, or "games," each of which teach one or more components of the Method of Difference. The first phase developed the Method of Difference. The first phase developed the Method of Difference behavior in a very deliberate fashion, and the second, using varied tasks, was intended to transfer the behavior under the control of abstract, schematic attributes of tasks and problems.

The first grade children used for the experiment were divided into two groups, only one of which received the programmed instruction. Details of the method, using live teachers, are given. Among the results the most striking was the attainment of positive, non-specific transfer of training. The results show children can acquire, retain and transfer rather complex problem-solving behavior, which is generally acquired at fourteen or fifteen. In that this program involves concrete materials and devices, it is consistent with the stage of intellectual development of six and seven year old children. Although children cannot conceptualize the systems they use, skills can be taught of considerable value, particularly in "new math" and science curricula. The results suggest that the presumption on which the project was based - that reasoning and problem solving capabilities consist of repertoires of behavior modifiable in accordance with principles of learning, instead of inaccessible structures and processes - can be the basis of feasible instruction.

ERIC No.
18. SOURCE OF ABSTRACT

C. FUNDING DATA

19. SUPPORTING ORGANIZATION
20. DIVISION OR DEPT.
21. CONTRACT NO.
PROJECT NO.
22. TOTAL COST OF PROJECT
23. DURATION OR PROJECT

U.S. Department of Health, Education, and Welfare
Office of Education

3

ERIC Form 1

NO. 1790A

ANDERSON

U. OF ILL.

ED001232

# SHAPING LOGICAL BEHAVIOR IN SIX- AND SEVEN-YEAR OLDS

COOPERATIVE RESEARCH PROJECT NO. 1790A

RICHARD C. ANDERSON

UNIVERSITY OF ILLINOIS

REPRINTED BY PERMISSION
U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN PRINTED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

# SHAPING LOGICAL BEHAVIOR IN SIX- AND SEVEN-YEAR OLDS

Cooperative Research Project No. 1790A

Richard C. Anderson

University of Illinois

Urbana, Illinois

1964

The research reported herein was supported by the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education and Welfare.

TABLE OF CONTENTS

Section 1. Introduction .......................................... 1

Frame of reference .......................................... 2

Description of approach .......................................... 5

Section 2. Analysis of Problem-Solving Strategies .......................................... 8

Section 3. A Procedure for Teaching the Method of Difference .......................................... 40

Overview of training procedure .......................................... 43

Details of training procedure .......................................... 44

Vanishing the control of task characteristics .......................................... 57

Section 4. The Measurement of Problem-Solving Behavior .......................................... 60

Presentation of problems .......................................... 64

Instructions .......................................... 65

Content analysis of problem-solving protocols .......................................... 73

Three summary measures of problem-solving performance .......................................... 76

Reliability of problem-solving measures .......................................... 79

Interrater reliability .......................................... 79

Shapes of the distributions of measures .......................................... 80

Internal-consistency reliability .......................................... 82

Other relationships .......................................... 88

Summary .......................................... 95

Section 5. Design of the Experiment .............................. 97

Sample .............................................................. 97
Training procedure ............................................ 98
Procedure for presenting problems ........................ 99
Statistical analysis .......................................... 102

Section 6. Results and Discussion ............................. 103

Acquisition of problem-solving behavior .................... 103
Effects of sex ..................................................... 112
Retention of problem-solving behavior ...................... 112
Transfer of problem-solving behavior ....................... 121
Effects of order .................................................. 135
Effect of training on aptitude factor scores ............... 135

Section 7. Summary and Conclusions ............................. 140

The experiment .................................................. 142
Results ............................................................. 147
Implications ....................................................... 149

References .......................................................... 156

## LIST OF TABLES

Table 1, Section 4. Interrater Reliability Coefficients for Three Summary Measures of Problem-Solving ......... 81

Table 2, Section 4. Kurtosis and Skewness of Distributions of Number of Unnecessary Trials ................ 83

Table 3, Section 4. Kurtosis and Skewness of Distributions of Invalid Conclusions ........................ 84

Table 4, Section 4. Kurtosis and Skewness of Distributions of the Square Root of Invalid Conclusions ...... 85

Table 5, Section 4. Correlations Between the Same Measure Scored from Two Similar Problems Involving the Same Task .................................. 87

Table 6, Section 4. Correlations Among Numbers of Invalid Conclusions by Task ........................ 89

Table 7, Section 4. Correlations Among Numbers of Unnecessary Trials by Task ............................. 90

Table 8, Section 4. Correlations Among Numbers of Solutions by Task ......................................... 91

Table 9, Section 4. Correlations Among Measures of Problem Solving Within Task .................................. 92

Table 10, Section 4. Correlations of Mental Age with Measures of Problem Solving by Task ..................... 93

Table 11, Section 6. Cumulative Number of Subjects Reaching Criterion of Mastery on Training Tasks Within 27 20-Minute Training Sessions .............. 104

Table 12, Section 6. Number of Subjects who Failed to Reach Criterion of Mastery on Training Tasks ...... 105

Table 13, Section 6. Median Problems to Criterion on Training Tasks ........................................ 107

Table 14, Section 6. Median Problems to Criterion on Three Selected Training Tasks by Order of Presentation ................................ 111

Table 15, Section 6. Number of Solutions to Cowboy Problems ..... 113

Table 16, Section 6. Number of Solutions to Pegboard Problems ... 114

Table 17, Section 6. Means and Standard Deviations of Numbers of Unnecessary Trials on Retention Problems ... 115

Table 18, Section 6. Analysis of Variance of Unnecessary Trials on Retention Problems ..................... 116

Table 19, Section 6. Means and Standard Deviations of Numbers of Invalid Conclusions on Retention Problems .. 118

Table 20, Section 6. Analysis of Variance of Invalid Conclusions on Retention Problems ..................... 119

Table 21, Section 6. Analysis of Variance of Solutions to Transfer Problems ................................ 123

Table 22, Section 6. Number of Solutions to Screw Problems ...... 124

Table 23, Section 6. Number of Solutions to Suburbia Problems ... 125

Table 24, Section 6. Number of Solutions to a Pendulum Problem .. 126

Table 25, Section 6. Number of Solutions to Chemical Problems ... 127

Table 26, Section 6. Number of Solutions to Pencil Problems ..... 128

Table 27, Section 6. Means and Standard Deviations of Numbers of Unnecessary Trials on Transfer Problems .... 129

Table 28, Section 6. Analysis of Variance of Unnecessary Trials on Transfer Problems ...................... 130

Table 29, Section 6. Means and Standard Deviations of Numbers of Invalid Conclusions on Transfer Problems ... 132

Table 30, Section 6. Analysis of Variance of Invalid Conclusions on Transfer Problems ...................... 133

Table 31, Section 6. Treatment by Order Analysis of Unnecessary Trials on Transfer Problems ................ 136

Table 32, Section 6. Treatment by Order Analysis of Invalid Conclusions on Transfer Problems ........... 137

Table 33, Section 6. Treatment by Order Analysis of Solutions on Transfer Problems .......................... 138

## LIST OF ILLUSTRATIONS

| | | |
|---|---|---|
| Figure 1, Section 2. | Card Array I | 11 |
| Figure 2, Section 2. | Illustration of an Ideal Method of Difference Solution to a Card Array Problem | 13 |
| Figure 3, Section 2. | Truth Functions | 15 |
| Figure 3A, Section 2. | Truth Tables for Functions | 15 |
| Figure 4, Section 2. | Schematic Representation of a Method of Difference Solution to a Problem from a Three-Variable System | 16 |
| Figure 5, Section 2. | Diagram of Strategy that Involves Processing of Positive and Negative Information, Memory for Instances, Conclusion Drawing Following Complete Selection of Instances, and Instance Selection and Progress Tracking Procedures from Conclusion Drawing | 25 |
| Figure 6, Section 2. | Diagram of a Strategy Involving Processing of Positive Information Only, Memory for Instances, Conclusion Drawing Following Complete Selection of Instances, and Instance Selection and Progress Tracking Procedures Separate from Conclusion Drawing | 26 |
| Figure 7, Section 2. | Diagram of a Strategy that Involves Processing of Positive Instances Only, Memory for Conclusions Drawn After Each Instance, Instance Selection Integrated with Conclusion Drawing, and Autonomous Progress Tracking | 28 |
| Figure 8, Section 2. | Diagram of a Strategy that Involves Processing of Negative Information only, Memory for Conclusions Drawn after each Instance, and Instance Selection and Progress Tracking Separate from Conclusion Drawing | 30 |

Figure 9, Section 2. Diagram of a Strategy that Involves Processing of Positive and Negative Instances, Memory for Conclusions Drawn after each Instance, Instance Selection and Progress Tracking Procedures Integrated with Conclusion Drawing .................................... 32

Figure 10, Section 2. Diagram of a Strategy that Involves Processing of Positive and Negative Instances, Memory for Conclusions Drawn after each Trial, Progress Tracking Integrated with Conclusion Drawing, and an Autonomous Instance Selection Procedure .......................... 34

Figure 11, Section 2. Diagram of a Strategy that Involves Processing of Positive and Negative Instances, Memory for Conclusions Drawn after each Instance, and Instance Selection and Progress Tracking Functions that are Autonomous from Conclusion Drawing .......................... 36

Figure 12, Section 4. Summary of Content Analysis Schema ........... 77

## ACKNOWLEDGEMENTS

I wish to thank Joffrey Nuney, Superintendent, Public Schools, East Brunswick, New Jersey for his cooperation in making the study possible. I wish to express my gratitude to the principals of the elementary schools in East Brunswick and, especially, the first grade teachers who graciously permitted their classes to be interrupted over a period of many months.

Peter Buermann, Richard Cummings, G. Herbert Fraunfelker, and Robert Vojack served as Research Assistants. What success the project has had is due in a large measure to the intelligence and energy they gave.

Last, but not least, I want to express my appreciation to Mrs. Gertrude Auger and Miss Bonnie Anderson, who typed reams of manuscript with skill and diligence.

Richard C. Anderson
University of Illinois, Urbana

# 1. INTRODUCTION

The project described in this report had two main goals. The first aim was to develop and evaluate a procedure for teaching children an inductive reasoning capability. Put simply, the goal was to teach children a classical method of experimental science, the strategy of varying each factor in succession while holding all other factors constant. As is detailed in the next section, a surprisingly large number of problems can be solved when this strategy is employed. Furthermore, the strategy as a whole, or elements of the strategy, may serve as parts or tactics in other strategies.

The strategy of varying one factor at a time while holding the rest constant is not ordinarily seen until fourteen to sixteen years of age. Inhelder and Piaget (1958, p. 62) argue that the strategy "involves the whole interpropositional combinatorial system" and that it cannot be mastered until the most advanced stage of cognitive development. Evidently, a procedure that successfully taught the strategy to children of six and seven years of age would have social significance.

The second goal of the project was to make a contribution to the understanding of complex human processes. The hope was, first, to make a more comprehensive analysis of a type of problem than had heretofore been done and, second, to make some improvements in the measurement of problem-solving behavior.

Frame of reference. The project herein described had more orientation of technology than of science. That is to say the project was chiefly concerned with inventing or creating a socially useful product -- namely, an instructional procedure -- rather than testing hypotheses or establishing empirical laws.

In his presidential address to the Division of Educational Psychology of the American Psychological Association, Donald Snygg (1963) argued that the stimulus-response orientation to psychology should be abandoned because, in his judgement, it has little implication for education. It is the writer's contention that basic research and theory seldom, if ever, have direct implications for practice and that statements of psychological laws are never of very much direct value to practical educators.

Consider one of the most widely agreed upon "laws" of behavioral science, the law of effect. The law of effect asserts that the strength of the learning of any response is related to the number of times this response has been reinforced. What is the response? What is a reinforcement? What should be done when a child simply doesn't give the desired response? These are questions which the teacher seriously attempting to apply the law of effect must certainly ask. Note that I am assuming that the teacher has a full understanding of the concepts "response" and "reinforcement." The questions that our hypothetical teacher would be compelled to raise are of a rather profound nature.

What he must have are realistic, working definitions of "response" and "reinforcement" which have reference to human beings studying reading, mathematics, history, or science. Even if generally satisfactory answers to these questions were forthcoming, one could not expect teachers explicitly to apply behavioral science principles.

It would take an extremely perceptive, quick-witted teacher instantaneously to classify a response as meeting or not meeting a criterion of success, take steps to elicit successful responses from students who fail, and plan the next step for those who succeed. The definition of a successful response necessarily changes as the subject matter changes and as the student progresses. No matter how skilled and dedicated the teacher -- particularly under the arrangement where the teacher must instruct 20-40 students an hour for 20-30 hours a week -- consistent, subtle application of behavioral principles seems virtually impossible. The more likely rendering of the law of effect by the teacher who takes his behavioral science seriously is, "I praise my students when they do good work," or, still further from the target, "I maintain a warm, supportive climate in my classes." There is no good reason for believing that any human being, depending upon on the spot improvisation, can precisely apply behavioral principles, at least in the enormous variety of contexts presented by conventional school arrangements. If education continues to depend upon extemporization by teachers to bridge the gap between research and practice, principles of learning are bound to be reduced to trivial generalities.

What is needed is a technology to bridge the gap between research and practice. Engineering plays a critical role in the application of scientific principles in any area. Take the development of rockets, for example. One could not deduce a rocket or even a blueprint for a rocket solely from the principles of physics and chemistry. Physical laws serve as guidelines in the design and construction of a rocket, but there are countless decisions that must be made for which scientific principles are of little or no help. A rocket could not be built without the use of rules-of-thumb and a good deal of trial-and-error. The atomic bomb is often mentioned as though nuclear fission were a deductive certainty. Actually, the prediction that the atomic bomb would explode was "an enormous extrapolation" from the physics of 1940. No one could be sure that the phenomenon known as critical mass was an experimental reality until atomic piles and bombs were built and tried (Conant, 1952, p. 12).

Engineering will certainly be of great importance with regard to social invention. The application of principles from even the relative advanced sciences always involves considerable trial-and-error. Unquestionably, the same will be true of behavioral science. The "degree of empiricism" in the behavioral sciences is, and probably always will be, very high.

It is perhaps fair to say that educators have held romantic notions about the power of basic research, theory, and the "scientific process." On the other hand, it is unquestionably true that much applied educational research is characterized by a very low level of empiricism, the trying out of all manner of schemes and *ad hoc* ideas, the collection of mountains of trivial data.

It is decidedly not the intent to support the latter practices. What is needed is a systematic, programatic technology that maintains intimate ties with basic research in the behavioral sciences, but which is not entirely subservient to the goals and rituals of science. Measurement and test construction constitute one area in which a fairly substantial technology has developed. Programed instruction represents another area in which a technology is rapidly emerging. It is with these areas that the project being described is identified.

*Description of approach.* Problem solving has been analysed from many points of view and at many levels of discourse. The task of constructing a viable program for teaching problem-solving behavior seems to require analysis at at least two levels.

The point of the first level is to develop a *prescription*. The terminal behavior is defined and a route -- or alternative routes -- to the desired terminal state is mapped. A special debt is owed to Bruner, Goodnow, and Austin, whose provocative book, *A Study of Thinking*, influenced the present effort both conceptually and methodologically. The functional approach that they developed is especially suited to analysis of problem-solving at a molar, prescriptive level. Such concepts as efficiency, risk regulation, cognitive strain are useful in conceiving ideal strategies.

To the extent possible, information theory (c.f., Hovland, 1952) and particularly symbolic logic have been employed in developing the prescription. Piaget (Piaget, 1953; Inhelder and Piaget, 1958; Flavell, 1963) served as a model in the latter regard. Flow diagrams of strategies that will yield the terminal behavior were constructed. They are useful in formalizing an analysis and checking it for completeness, consistency and sufficiency.

The second level of analysis is behavioral. A problem-solving strategy can be conceived as a sequence of behavior or a behavioral chain (c.f., Bruner, et al, 1956, p. 54). Strategies are under stimulus control of two types. The first type of control is exercised by task or problem cues. Strategies are also controlled by response produced stimulation, including intraverbals (Skinner, 1957). It is the latter source of control that gives problem-solving behavior its coordinated and purposeful character.

The proof of a behavioral analysis is synthesis. If one can synthesize the prescribed behavior, then the analysis is at least satisfactory, though almost certainly not unique. The general question is how do you get children to perform a given sequence of problem-solving behavior. The specific questions for each part of the sequence are as follows. Exactly what behavior is required and how do you evoke the behavior or, if necessary, shape the behavior? Precisely what stimuli should control the behavior and how is this control to be arranged? Answers to the questions of what behavior and what stimulus control are judged against the prescription developed in the first phase of analysis. Answers to the how questions are formulated on the basis of the emerging instructional technology as well as the basic research literature.

## 2. ANALYSIS OF PROBLEM-SOLVING STRATEGIES

The goal of this research and development enterprise is to teach children the problem-solving strategy in which each factor is varied in succession while all other factors are held constant. The first item on the agenda is to determine precisely what is involved in this strategy, what are the components, and how are they to be put together. In short, the problem is to form a prescription.

There are several ways to proceed in analyzing an inductive reasoning capability. One approach would be to employ dimensions established by factor analysis as conceptual guidelines. Instructional goals and subgoals could be couched in terms of factors. L. L. Thurstone (1948) believed that "primary mental abilities could be improved by training." A set of exercises designed for children in the primary grades, called the Learning to Think Series, was published under the authorship of Thelma Thurstone (1949). Indications are that the exercises have not been employed in experimental investigations nor used extensively in schools. Torrance has been concerned with learning to "think creatively." He and his associates (Meyers, 1961) have developed exercises designed to improve Ideational Fluency and Originality. Holland is currently attempting to develop a sequential, small step program to improve Spatial Relations. He has not yet reported his work.

The factor analysis approach is less than satisfactory for dealing with complex patterns of behavior. A task may be named, that is classified in a slot in a taxonomy, but there the analysis stops. If the goal is to obtain a prescription itemizing what should be taught if a complex task is to be mastered, the result is likely to be a recipe of the form two parts Verbal Comprehension, one part General Reasoning, and a dash of Number Facility.

A way of analyzing problem-solving is needed which allows one to cope with intratask relationships and sequential dependencies. To this end, a functional and logical analysis was employed. The technique is to develop ideal or idealized strategies after the manner of Bruner, _et_ _al_., (1956, pp. 83-90). These strategies can be conceptualized in terms of a sequence of operations. Their generality, limitations, efficiency, completeness, and consistency can be judged.

Let us now begin to apply these conceptions to the method of varying one factor at a time while holding all other factors constant. This strategy was called the Method of Difference by John Stuart Mill, who was among the first to analyze inductive methods rigorously. Stebbing (1888, p. 120) wrote the following about the Method of Difference in _Analysis_ _of_ _Mr_. _Mill's_ _System_ _of_ _Logic_:

> "_Canon_ -- If an instance in which the phenomenon in question occurs . and an instance in which it does not occur have every other circumstance in common save one, that one occurring only in the former, -- that circumstance in which

alone the two instances differ, in the effect or the cause or a necessary part of the cause, of the phenomenon.

The principle is that of comparing an instance of the occurrence of a phenomenon with a similar instance in which it does not occur, to discover in what they differ.

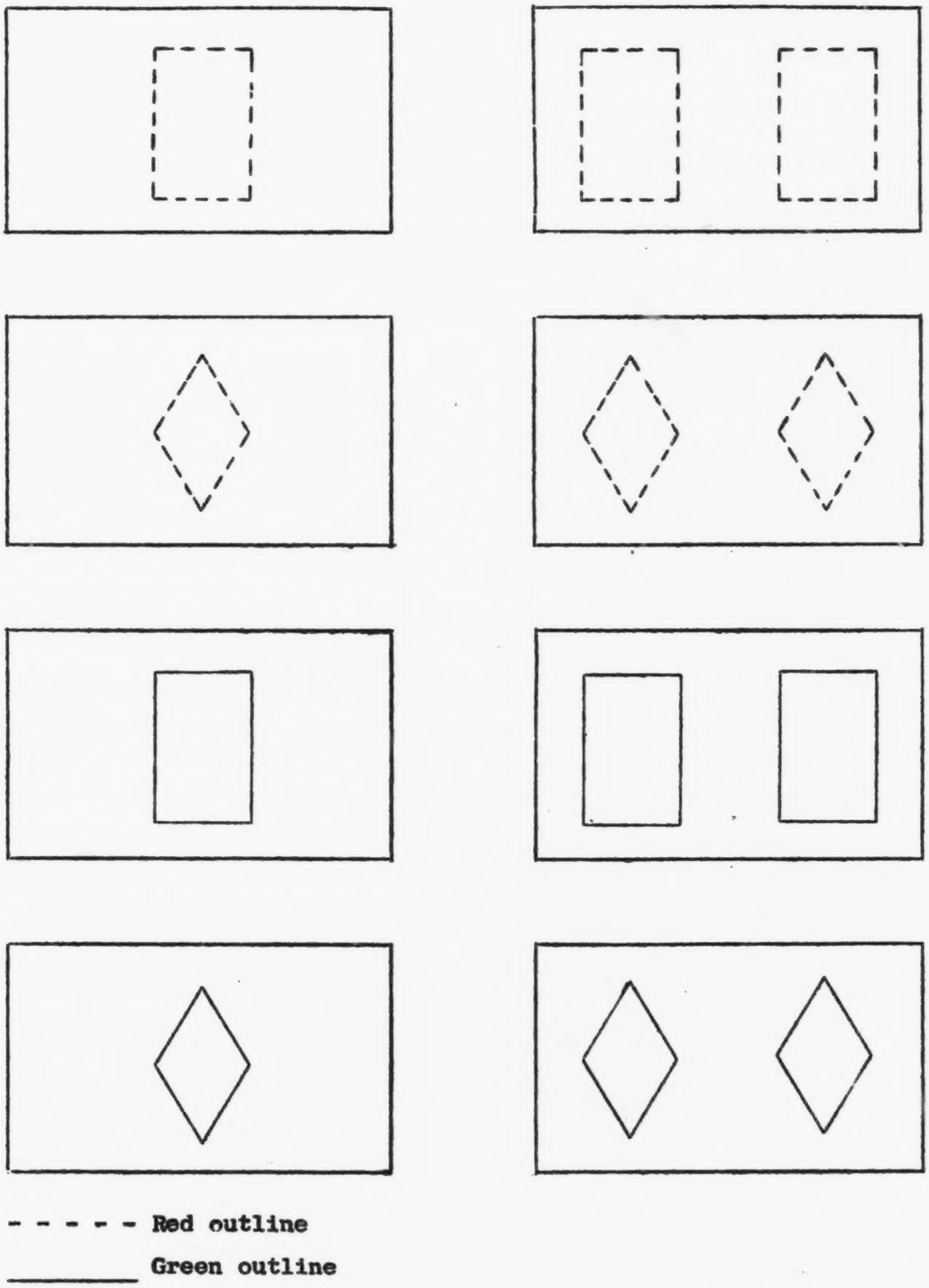
Remarks:--

1. This method is more particularly a method of artificial experiment (its ordinary use being to compare the condition of things before, with those after, an experiment), because --
2. It is commonly employed to determine the effects of given causes; and because --
3. The instances which it requires are rigid and definite -- they must be exactly alike, except that in one the phenomenon must be present and in the other absent.
4. If this method is inapplicable, it is usually because artificial experiment is impracticable.
5. It is the only method, of direct experience, by which laws of causation can be proved.
6. If the instances fulfill exactly the requirements of the Canon, this method is perfectly rigorous in its proof."

The Method of Difference will be illustrated with one of the tasks used in training, an array of eight cards. The array represents two values of each of three variables, as is shown in Figure 1. The values of the variables are coded as follows:

Figure 1

Card Array I


- - - - - Red outline
Green outline

| | Values | Code |
|---|---|---|
| Number | | |
| | One figure | 1 |
| | Two figures | 2 |
| Color | | |
| | Red | R |
| | Green | G |
| Shape | | |
| | Box | b |
| | Diamond | d |

An instance is described by listing the number, color, and shape codes in that order. Thus, the card showing two red diamonds is coded 2Rd. A "+" means that the instance shows the concept and a "-" means that the instance does not show the concept.

The instructions for the card array task are as follows. "I will think of a secret and point to a card that shows my secret. You try to figure out my secret by pointing to other cards. Each time you point to a card I will tell you if it shows my secret or not. As soon as you're sure you know, tell me the secret." The initial positive instance selected by the experimenter is called the "focus instance." Each card selected by the subject is called a "trial." Figure 2 contains an illustration of a Method of Difference solution to a card array problem.

The hallmark of the Method of Difference is that there is information available from every trial. When one variable at a time is changed, the information that the variable that was manipulated is irrelevant is available when the resulting instance is positive. When the

## Figure 2

### Illustration of an Ideal Method of Difference Solution to a Card Array Problem

| | |
|---|---|
| Focus instance: | 1Rb+ |
| First trial: | 2Rb+ |
| Second trial: | 2Gb- |
| Third trial: | 1Rd- |
| Conclusion: | Red box |

NOTE:--See text for interpretation of codes.

resulting instance is negative, the information that the state of the variable that is contained in the focus instance is part of the concept is available. Any method that involves changing more than one variable at a time may produce trials in which little or no information is available. The Method of Difference is the most efficient strategy there is for attaining conjunctive concepts, most efficient in the sense that it minimizes the maximum number of trials that will logically be required for concept attainment. A concept can always be attained by the Method of Difference in exactly as many trials as there are variables to be processed. Any other strategy will on the average (assuming a random series of concepts to be attained) require more trials than there are variables. The superior efficiency of the Method of Difference follows from the well-known theorem of information theory that the most efficient information processing procedure is the one that reduces the set of possible concepts by exactly half at each step. The Method of Difference is the only strategy for attaining conjunctive concepts that reduces the set of possible concepts by exactly half at each trial.

The Method of Difference can be depicted schematically. Let "p," "q," and "r" be the assertions of given values of the variables $X_1$, $X_2$, and $X_3$ respectively. The statement of an instance is a conjunction containing as elements an assertion of one value (either the given value or not the given value) of each variable in the system. Let "+" represent the statement, when conjoined with an instance, that the

Figure 3

Truth Functions

| Function | Symbol | Rough English equivalent |
|---|---|---|
| Negation | - | Not |
| Conjunction | .* | And |
| Disjunction | $\vee$ | Or (inclusive from) |
| Conditional | $\supset$ | If then |
| Biconditional | $\equiv$ | If and only if |

*Also, no punctuation.

Figure 3A

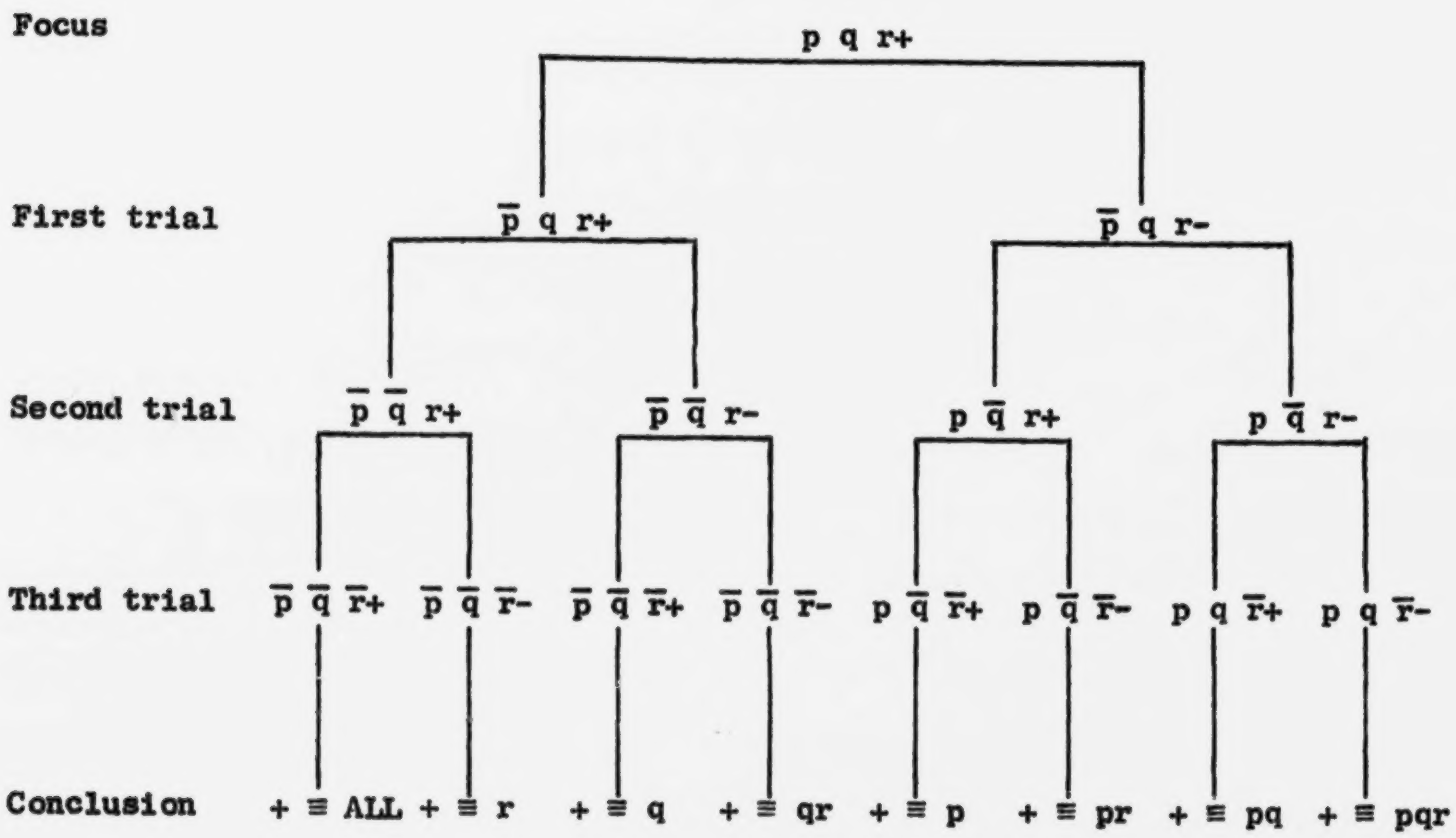
Truth Tables for Functions

| Functions | | | | | Statements | |
|---|---|---|---|---|---|---|
| $\bar{p}^*$ | $p \equiv q$ | $p \supset q$ | $p \vee q$ | $p . q$ | p | q |
| | T | T | T | T | T | T |
| | F | F | T | F | T | F |
| | F | T | T | F | F | T |
| | T | T | F | F | F | F |

| $\bar{p}$ | p |
|---|---|
| F | T |
| T | F |

Figure 4

Schematic Representation of a Method of Difference Solution to a Problem from a Three-Variable System


Focus
$p\ q\ r+$
First trial
$\bar{p}\ q\ r+$
$\bar{p}\ q\ r-$
Second trial
$\bar{p}\ \bar{q}\ r+$
$\bar{p}\ \bar{q}\ r-$
$p\ \bar{q}\ r+$
$p\ \bar{q}\ r-$
Third trial
$\bar{p}\ \bar{q}\ \bar{r}+$
$\bar{p}\ \bar{q}\ \bar{r}-$
$\bar{p}\ \bar{q}\ \bar{r}+$
$\bar{p}\ \bar{q}\ \bar{r}-$
$p\ \bar{q}\ \bar{r}+$
$p\ \bar{q}\ \bar{r}-$
$p\ q\ \bar{r}+$
$p\ q\ \bar{r}-$
Conclusion
$+ \equiv$ ALL
$+ \equiv r$
$+ \equiv q$
$+ \equiv qr$
$+ \equiv p$
$+ \equiv pr$
$+ \equiv pq$
$+ \equiv pqr$

instance shows the concept, and let "-" represent the statement, when conjoined with an instance, that the instance does not show the concept.[1] Figure 4 contains a schematic representation of the Method of Difference in the system which has just been defined.[2]

Strictly speaking, the illustration of the Method of Difference presented in Figure 2 would have to be rewritten before a rigorous logical analysis could be undertaken. Since in the card array that was described a card is either red or green and there are no other possibilities, it is "obvious" that if the card two-red-box shows the secret (i.e., 2Rb+) and that if the card two-green-box does not show the secret (i.e., 2Gb-), then *red* is part of the secret. To be precise, however, 2Rb+ and 2Gb- do *not* imply red is part of the secret. If two-green-box (2Gb) is rewritten two-not-red-box ($2\overline{R}b$) then the statement of the instances does imply the conclusion in strict logical terms. Schematically, the assertion is as follows:

$$(2Rb+ \vee 2\overline{R}b-) \supset (+ \supset R)$$

---

[1] In other words, plus (+) is defined as equivalent to not-minus ($\overline{-}$) and minus (-) is defined as equivalent to not-plus ($\overline{+}$).

[2] Figure 4 does not contain all of the acceptable instances that might appear in a Method of Difference solution. After a variable has been found irrelevant (that is after two values of the variable appear in otherwise identical positive instances) either the given value or its negation may appear in subsequent instances without affecting the conclusion. Thus, in the illustration in Figure 2, the instances occurring at the second and third trial could contain either 1 or 2.

Whereas it may seem trivial to insist upon not-red instead of green in this case, the necessity for the distinction can be more readily seen in a case where there were four colors involved instead of two. The Method of Difference can be applied to attain conjunctive concepts regardless of the number of values the variables take on. This assertion follows from the definition of an instance as a conjunction containing as elements assertions of exactly one value of each variable in the system. Stated schematically, the assumption is as follows. Let $P_1$, $P_2$, $P_3$ .... $P_j$ be statements of the values of the variable X. It is assumed that

$$\overline{P}_1 \equiv P_2 \vee P_3 \cdots\cdots \vee P_j$$

or, equivalently, that

$$P_1 \equiv \overline{P}_2\ \overline{P}_3 \cdots\cdots \text{ and } \overline{P}_j.$$

As will be seen later, children who behave in a facile manner with systems in which no variable takes on more than two values frequently encounter difficulty with systems in which one or more variables take on three or more values until they have learned these equivalences.

Given any focus instance -- such as 1Rb+, the focus instance in the illustration in Figure 2 -- a schematic "tree" for the card array task could be constructed that is in one-to-one correspondence with the schematic "tree" in Figure 4. By the rule that substitution of schemata for schemata preserves validity, the schematic "tree" envisioned for the card array task that begins with the focus instance 1Rb+ may be altered by substituting p for each occurrence of l, q for each occurrence of R and r for each occurrence of b. Thus, it is seen that the schema in Figure 4 is a perfect representation of the card array task, indeed, a perfect representation of any three-variable conjunctive system. A representative schematic "tree" could be constructed for a system of any size. By size is meant the number of variables in the system for the number of possible concepts and the complete specification of the schematic tree is a function solely of the number of variables.[3]

What can be said about the relationship between conjunctive systems of different sizes? A many-one substitution (of the form p for each occurrence of q, r, ....) can be made in the schematic tree of a system which, after the deletion of redundant and inconsistent terms, is identical with the schematic tree for any given smaller system. However, a one-many substitution (qrs .... for each occurrence of p) in a schematic tree will not produce an enlarged tree that is in one-to-one correspondence with the tree for any given larger system. There is, in other words, a homorphic rather than an isomorphic relationship among systems of different sizes. It is not known whether this fact has any psychological significance. If it does, then one might expect that training on systems of a given size would transfer more readily to smaller than to larger systems.

---

[3] Let $N_1$, $N_2$, $N_3$ ... $N_k$ represent the number of values of each of k variables. The number of instances in a closed, conjunctive system is given by the expression $\prod_{i=1}^{k} N_i$. The number of possible concepts before the presentation of a focus instance is $\prod_{i=1}^{k} (N_i + 1)$. The number of possible concepts following the presentation of a focus instance is $2^k$.

It is a feature of the representative schematic "tree" for a conjunctive system that a disjunction of the instances actually observed always implies the appropriate conclusion and no other. For example, from the illustration in Figure 2, it can be verified that:

$$(1Rb+ \vee \overline{1}Rb+ \vee \overline{1}\overline{R}b- \vee 1R\overline{b}-) \supset (+ \equiv Rb).$$

This expression represents one way of defining the Method of Difference.

Suppose that a subject were presented with a series of card array problems of the sort previously described. Suppose further that in every case the subject states the correct conclusion after exactly three trials (it always being the case that the statement of the instances implies the conclusion and no other). It could in truth be asserted that this subject "knows" the Method of Difference, and that he is "using" or "applying" the method. However, this is an imprecise manner of speaking that contains misleading surplus meaning. Herein lies an epistemological confusion that has long plagued educators and psychologists, particularly cognitive psychologists.

When it is asserted that a person "knows" a rule there is the implication that the rule exists "in the head" in verbal form. If it is observed that the person cannot state the rule, though he performs well in a situation to which the rule is applicable, it is said that the rule exists in intuitive or non-verbal form and it is still frequently asserted that the person is "applying the rule." The person who is said to show an "intuitive comprehension" of a rule is tacitly ascribed all of the potency of the verbal form, except that, because of some communication deficiency, he can't state the rule. When, on the other hand, a person exhibits good "rule stating behavior" but defective performance, it is frequently asserted that he is "mouthing words," that he doesn't "really know" the rule.

All that is actually observed is certain behavior (in the context of this discussion either rule-stating behavior or problem-solving behavior) in the presence of certain stimuli. It may be that there is a correspondence, perhaps a one-to-one correspondence, between the stimuli-response sequence that has been observed and a rule or set of rules. In any event, the most ambitious statement that can ever be made about what a person knows is that his overt behavior is isomorphic with a set of rules. What is more, according to the philosopher Quine (1960) it is possible only to note correspondences with a pre-conceived set of rules, so, if he is correct, it is impossible in any simple sense to induce the rules with which behavior corresponds.

If a rock is thrown into the air, there is a one-to-one correspondence between a set of points in the air that mark the flight of the rock and the set of points on a graph that satisfy some equation. It would not be said that the rock was "applying" the equation nor that there was a mechanism in the rock programed according to the equation. The state of physics being what it is, these ascriptions are not likely to be made today. Similar ascriptions are still being made in psychology and education. Perhaps the surplus meaning intrinsic in such ascription is one explanation of the rampant optimism which is current, regarding "massive generalized transfer of training" and long-term retention, optimism that far exceeds what could actually be predicted on the basis of empirical studies.

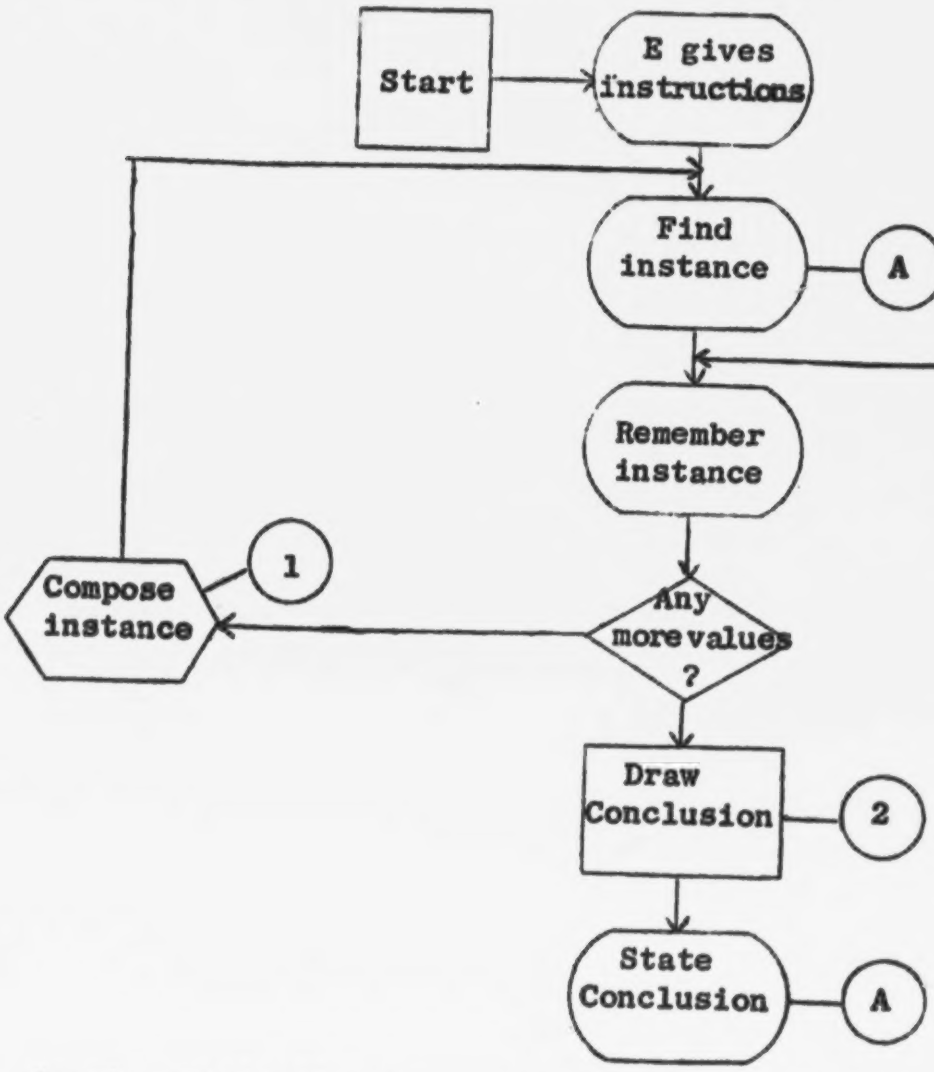
The instructional goal can now be precisely defined. The goal is concept-attainment behavior within any small, well-defined, conjunctive system that is isomorphic with the Method of Difference. By "well-defined" is meant capabilities of the subject, either entering behavior or behavior that can easily be acquired, with regard to the variables, the values which the variables take on, the instances, and the set of possible conjunctive concepts. A system is well-defined for a subject when he can name the variables (e.g., answer the question "In what ways are these things different?), name the values of each variable (e.g., answer the question "What are the different colors these cards show?), find or create the exemplifying instances given the name of any conjunctive concept (e.g., respond to the request "Find the compound, pinate leaves."), and name the conjunctive concept given any complete set of exemplifying instances (e.g., respond to the question "These cards (1Rb, 1Rd) show the secret; what is the secret?"). The aim is concept-attainment behavior that, for any problem, consists of a sequence of exactly k+1 overt responses, where k is the number of variables in the system from which the problem was selected. There are to be k choices of instances and a conclusion, such that a disjunction of the instances always implies the conclusion and no other.

There are an indefinite number of strategies that could produce the goal behavior described above. So, the Method of Difference is not a single strategy but a group of strategies. Consider now some of the properties of strategies that could yield the terminal concept attainment behavior. The strategy must include behavior that satisfactorily serves several functions, including instance selection, memory, conclusion drawing, and keeping track of progress through the problem. All instances may be remembered, or only positive instances, or only negative instances. Updated conclusions of several sorts may be remembered in addition to or instead of instances, or remaining possible concepts may be remembered. Conclusion drawing may begin after the first instance or not until all instances have been selected. Conclusions may deal with common elements, necessary values, irrelevant variables, possible concepts, or any combination of these. Instance selection may follow from otherwise useful memory or conclusion drawing functions, or there may be a relatively autonomous instance selection algorithm. Likewise, progress through the problem may be charted from other functions or there may be a separate index function.

Figures 5 through 11 contain diagrams of strategies illustrating different memory, instance selection, conclusion drawing, and progress tracking procedures. Which, if any, of these diagrams or other diagrams that could be constructed map the strategies that real people do employ, could learn to employ, or should learn to employ remains to be seen. Insofar as these strategies are or could be representations of reality, differential predictions about behavior may be made on a speculative, a priori basis. Some strategies impose a greater memory burden or greater computational strain than others, and are, hence, more prone to failure, particularly under limiting conditions. Some strategies will produce characteristic errors upon the incursion of certain disturbances. There will be differences in total time elapsed between the presentation of the focus instance and the statement of the conclusion, depending upon the strategy. There will also be differences in latency from response to response. There is little hope, or so it seems to the present author, of making fine discriminations among the problem-solving behaviors of persons "as they are." Presumably, ones current behavior is a function of ones history as well as current stimulation. What is needed to map the strategy of a person is a knowledge of his history. The only satisfactory method of filling this need is to teach the strategy, build the history in which one is interested. The strategies of people who exhibit the goal behavior without training, by "doin'

Figure 5

Diagram of a strategy that involves processing of positive and negative information, memory for instances, conclusion drawing following complete selection of instances, and instance selection and progress tracking procedures separate from conclusion drawing:


Start
E gives instructions
Find instance
A
E says + or -
Remember instance
Compose instance
1
Any more values ?
Draw Conclusion
2
State Conclusion
A

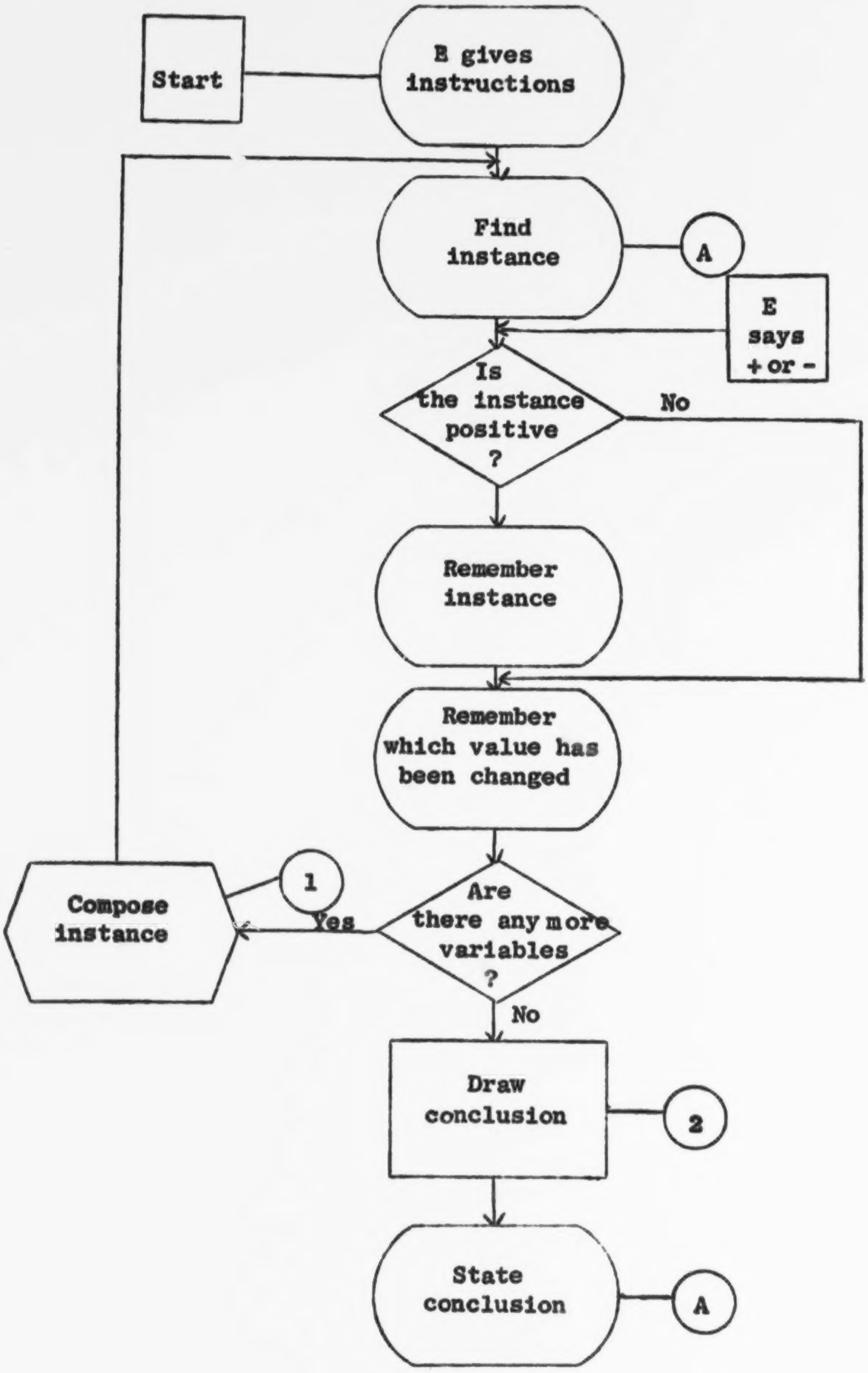
A Overt, publically observable behavior.

1 Instance composed according to a rule such as this: Next instance = a values of the last positive instance with the ith value changed (negated), where i = some value not previously changed.

2 There are several possible conclusion drawing algorithms in this case such as: Conclusion = the conjunction of each value that appears in a positive instance and whose negation appears in a negative instance otherwise the same as the positive instance.

Figure 6

Diagram of a strategy involving processing of positive information only, memory for instances, conclusion drawing following complete selection of instances, and instance selection and progress tracking procedures separate from conclusion drawing.


Start
E gives instructions
Find instance
A
E says + or -
Is the instance positive ?
No
Remember instance
Remember which value has been changed
Compose instance
1
Yes
Are there any more variables ?
No
Draw conclusion
2
State conclusion
A

(continued on next page)

Figure 6 (continued)

Ⓐ Overt, publicly observable behavior.

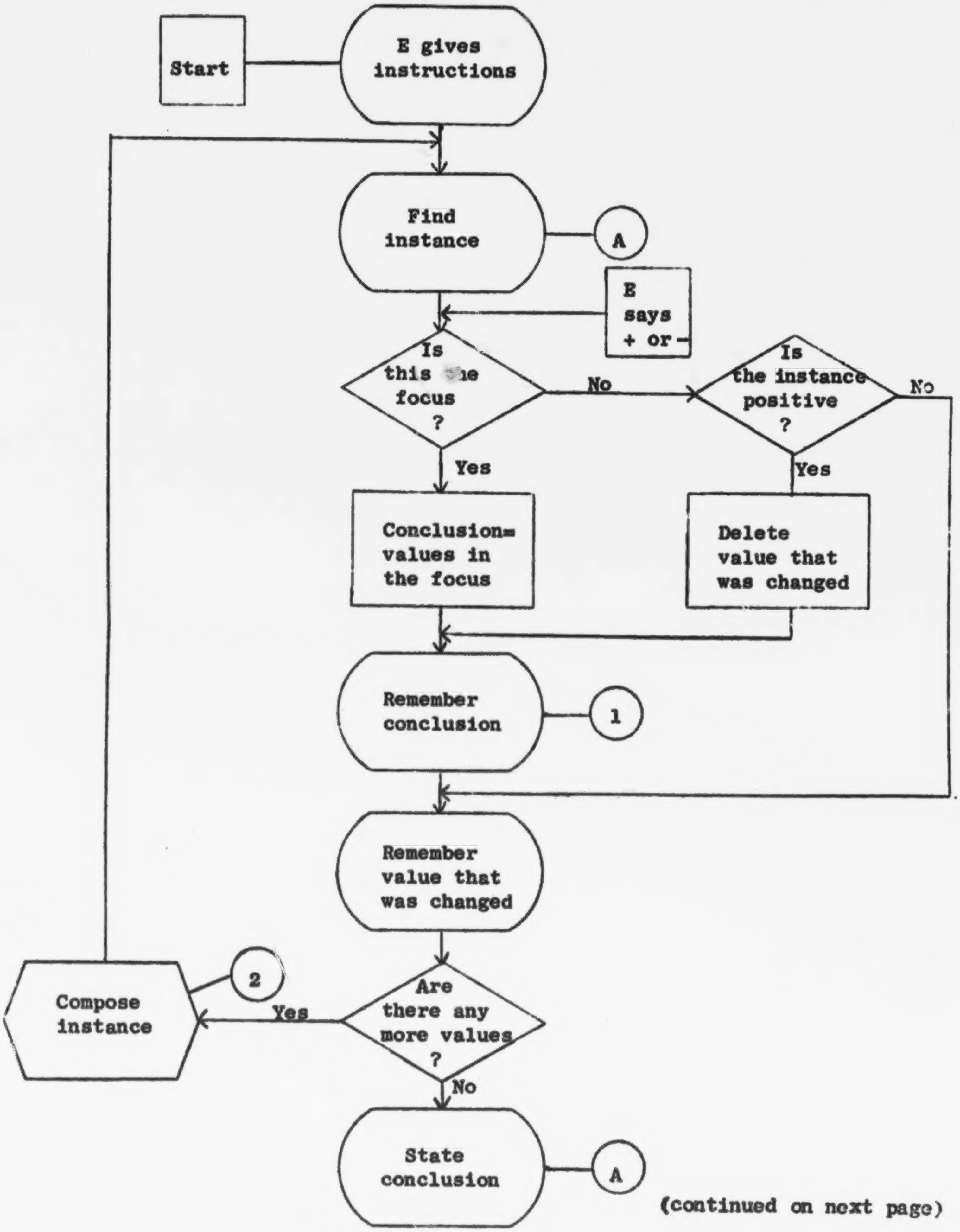
① Instances composed in the following manner:

Next instance = a) values occurring in the last positive instance with b) the ith value changed (negated) where i = some value not previously changed.

② Conclusion = values common to *all* positive instances.

Figure 7

Diagram of a strategy that involves processing of positive instances only, memory for conclusions drawn after each instance, instance selection integrated with conclusion drawing, and autonomous progress tracking:


Start
E gives instructions
Find instance
A
E says + or -
Is this the focus ?
No
Yes
Is the instance positive ?
No
Yes
Conclusion= values in the focus
Delete value that was changed
Remember conclusion
1
Remember value that was changed
Are there any more values ?
Yes
No
Compose instance
2
State conclusion
A

(continued on next page)

Figure 7 (continued)

Ⓐ Overt, publicly observable behavior.

① Contents of memory after each instance of card array problem illustrated in Figure 2 as follows:

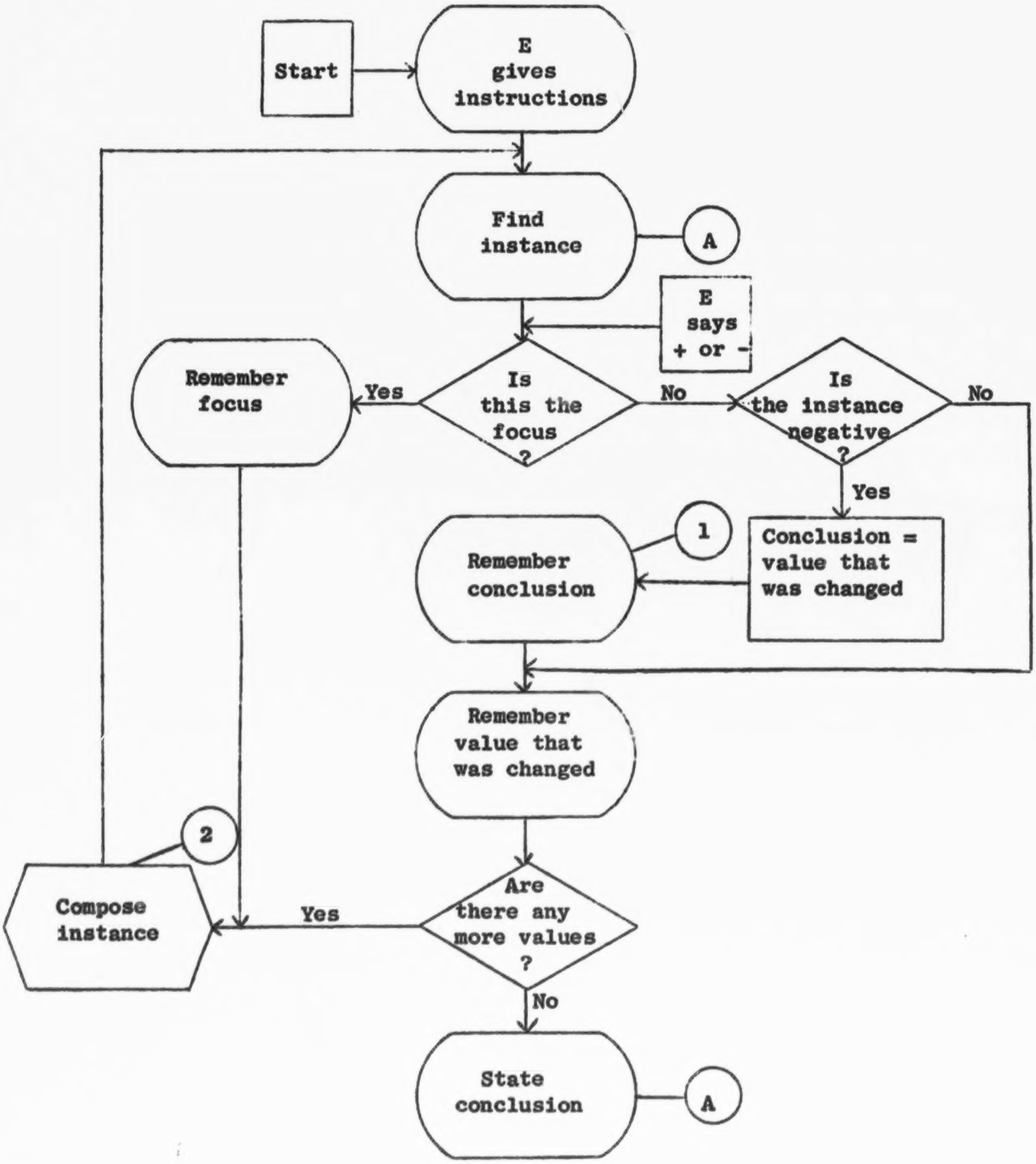
Focus instance: 1Rb
First trial: Rb
Second trial: Rb
Third trial: Rb

② Instances composed as follows:

Next instance = (a) values in current conclusion (b) with values not appearing in conclusion negated or not at random, and (c) with the ith value changed (negated) where i = a previously unchanged value.

Figure 8

Diagram of a strategy that involves processing of negative information only, memory for conclusions drawn after each instance, and instance selection and progress tracking separate from conclusion drawing:


Start
E gives instructions
Find instance
A
E says + or -
Remember focus
Yes
Is this the focus ?
No
Is the instance negative ?
No
Yes
Remember conclusion
1
Conclusion = value that was changed
Remember value that was changed
2
Compose instance
Yes
Are there any more values ?
No
State conclusion
A

(continued on next page)

Figure 8 (continued)

Ⓐ Overt, publicly observable behavior.

① Contents of memory after each instance of the card array problem illustrated in Figure 2 as follows:

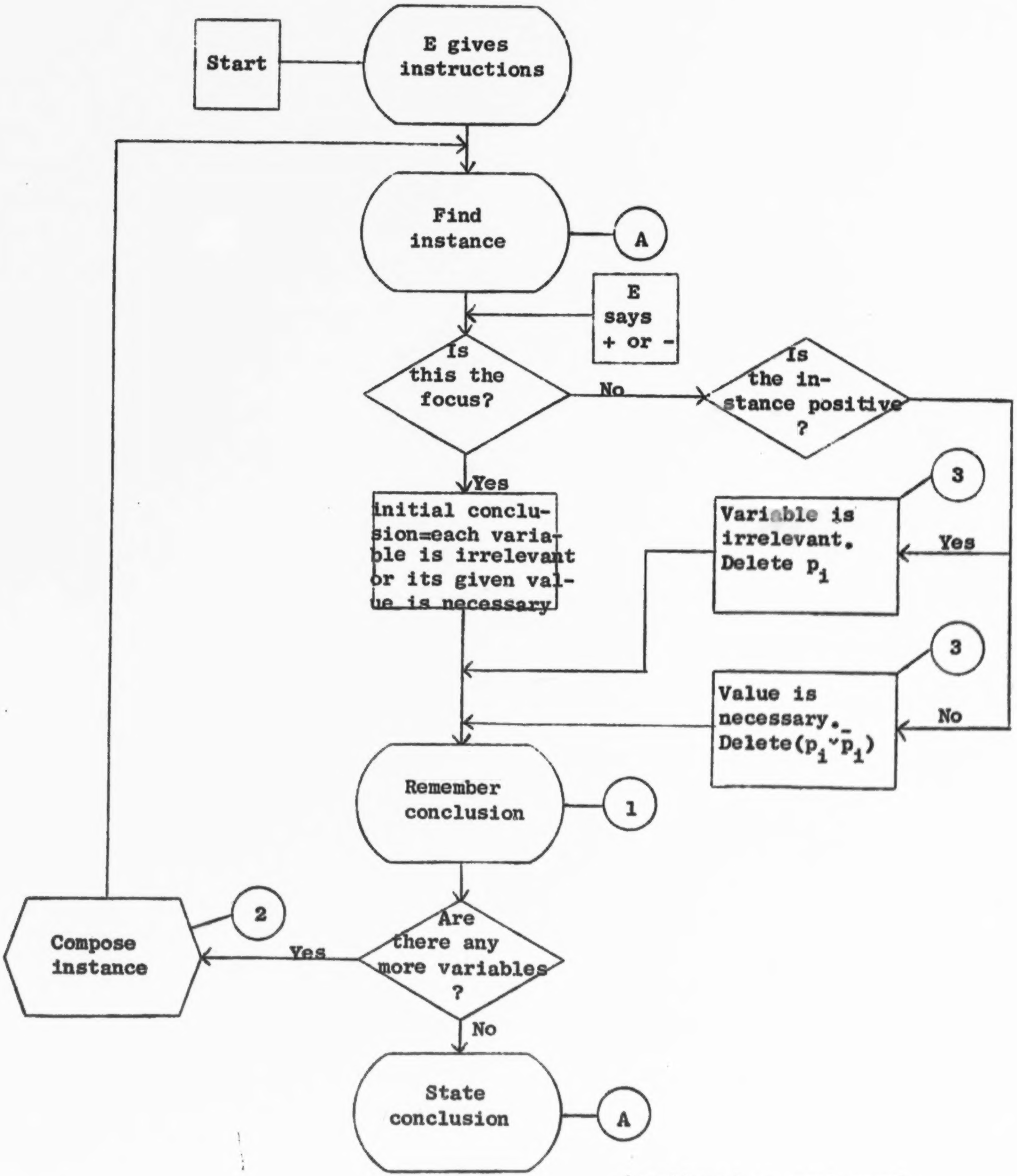
Focus instance: ———
First trial: ———
Second trial: R
Third trial: Rb

② Instances composed as follows:

Next instance = (a) values occurring in the focus with (b) the ith value changed (negated) where i = a previously unchanged value. This is a "conservative focusing" procedure (Bruner, _et al_, 1956, Pp. 87). The procedure would not yield the second trial instance of 2Gb- as occurred in the illustration in Figure 2 but rather 1Gb-.

Figure 9

Diagram of a strategy that involves processing of positive and negative instances, memory for conclusions drawn after each instance, instance selection and progress tracking procedures integrated with conclusion drawing:


Start
E gives instructions
Find instance
A
E says + or -
Is this the focus?
No
Is the instance positive ?
Yes
initial conclusion=each variable is irrelevant or its given value is necessary
3
Variable is irrelevant. Delete $p_i$
Yes
3
Value is necessary. Delete($p_i \vee \bar{p}_i$)
No
Remember conclusion
1
2
Compose instance
Yes
Are there any more variables ?
No
State conclusion
A

(continued on next page)

Figure 9 (continued)

**Ⓐ Overt, publicly observable behavior.**

**① Contents of memory after each instance of card array problem illustrated in Figure 2 as follows:**

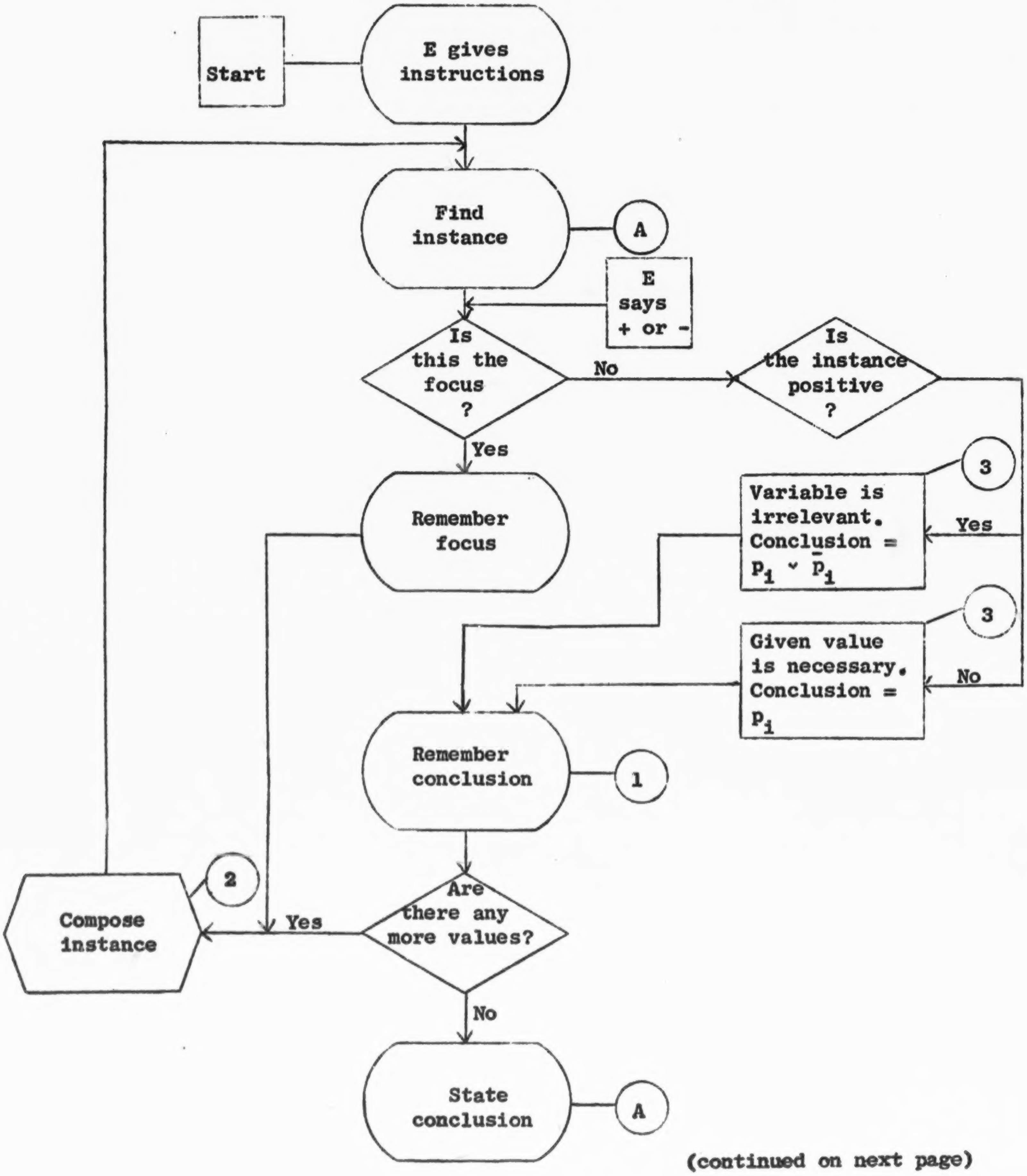
**Focus instance:** $1 \vee (1 \vee \bar{1})$. $R \vee (R \vee \bar{R})$. $b \vee (b \vee \bar{b})$
**First trial:** $1 \vee \bar{1}$. $R \vee (R \vee \bar{R})$. $b \vee (b \vee \bar{b})$
**Second trial:** $1 \vee \bar{1}$. $R$ . $b \vee (b \vee \bar{b})$
**Third trial:** $1 \vee \bar{1}$. $R$ $b$

**② Next instance = (a) necessary values as given in current conclusion (b) values of irrelevant variables negated or not at random (c) given values of untested variables except (d) the ith value is changed (negated) where i = a previously unchanged value (i.e. value of variable represented in current conclusion by $p_i \vee (p_i \vee \bar{p}_i)$).**

**③ Let $p_1$, $p_2$, $p_3$ ... $p_k$ represent given values of the variables $x_1$, $x_2$, $x_3$ ... $x_k$, respectively, rather than various values of a single variable, the definition used in the text.**

Figure 10

**Diagram of a strategy that involves processing of positive and negative instances, memory for conclusions drawn after each trial, progress tracking integrated with conclusion drawing, and an autonomous instance selection procedure:**


Start
E gives instructions
Find instance
A
E says + or -
Is this the focus ?
No
Is the instance positive ?
Yes
Remember focus
3
Variable is irrelevant. Conclusion = $P_i \vee \bar{P}_i$
Yes
3
Given value is necessary. Conclusion = $P_i$
No
Remember conclusion
1
2
Compose instance
Yes
Are there any more values?
No
State conclusion
A

(continued on next page)

Figure 10 (continued)

Ⓐ Overt, publicly observable behavior.

① Contents of memory after each instance of card array problem illustrated in Figure 2 as follows:

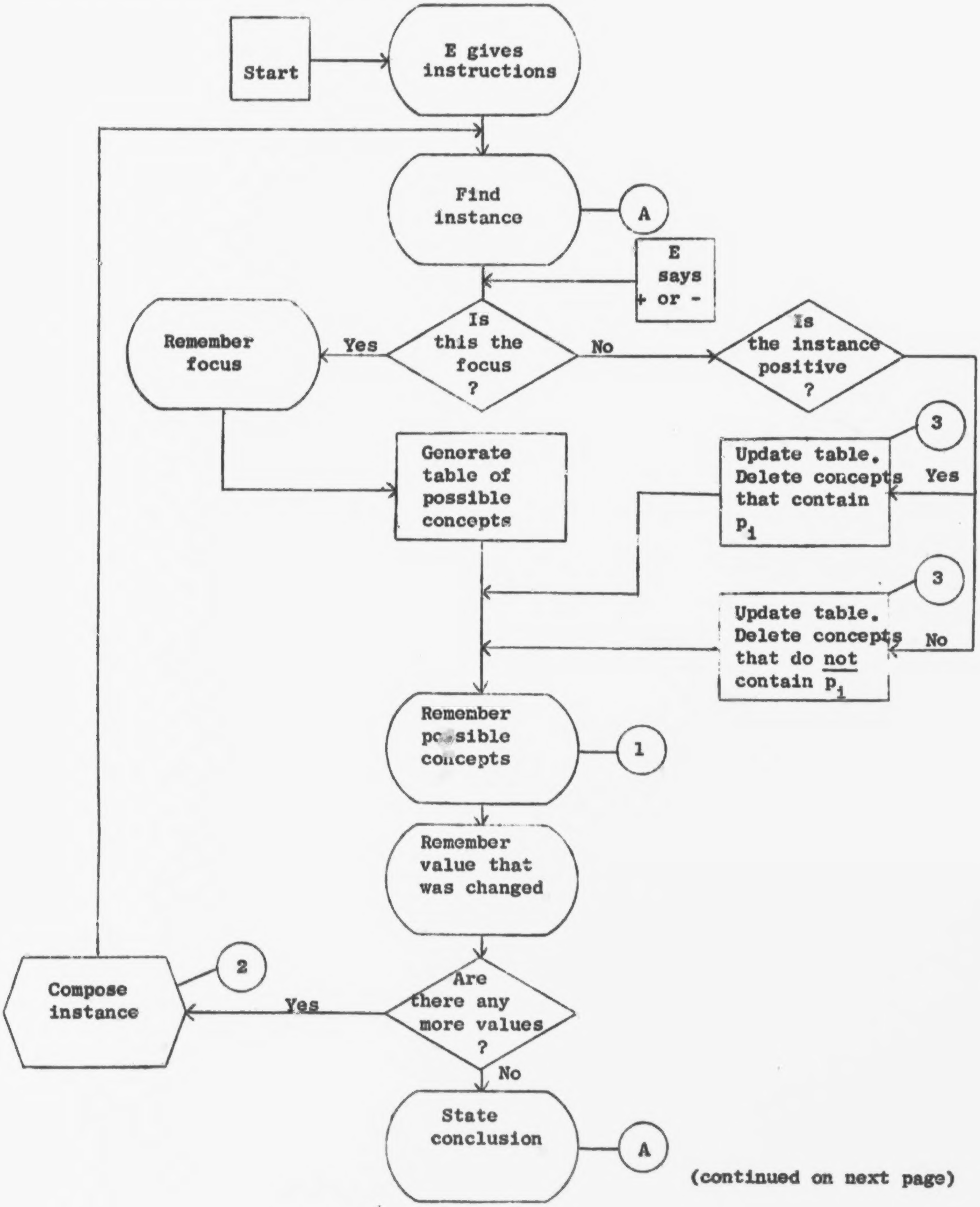
Focus instance: ———
First trial: $1 \vee \bar{1}$
Second trial: $1 \vee \bar{1}.\ R$
Third trial: $1 \vee \bar{1}.\ Rb$

② Next instance = (a) values occurring in the focus with (b) the ith value changed (negated) where i = a previously unchanged value. This is a conservative focusing procedure and the strategy as a whole resembles very closely the entire strategy that Bruner, et al, (1956) called "conservative focusing."

③ $p_1$, $p_2$, $p_3$ .... $p_k$ are given values (given in the focus) of the variables $x_1$, $x_2$, $x_3$ ... $x_k$, respectively.

Figure 11

**Diagram of a strategy that involves processing of positive and negative instance, memory for conclusions drawn after each instance, and instance selection and progress tracking functions that are autonomous from conclusion drawing:**


Start
E gives instructions
Find instance
A
E says + or -
Is this the focus ?
Yes
No
Remember focus
Is the instance positive ?
Generate table of possible concepts
3
Update table. Delete concepts that contain $p_i$
Yes
3
Update table. Delete concepts that do not contain $p_i$
No
Remember possible concepts
1
Remember value that was changed
2
Compose instance
Yes
Are there any more values ?
No
State conclusion
A

(continued on next page)

Figure 11 (continued)

Ⓐ Overt, publicly observable behavior.

① Contents of memory after each instance of the card array problem illustrated in Figure 2 as follows:

Focus instance: 1Rb, 1R, 1B, Rb, 1, R, b, All
First trial: Rb, R, b, All
Second trial: Rb, R
Third trial: Rb

② Next instance = (a) values accurring in the focus with (b) the ith value changed (negated), where i = a previously unchanged value.

③ $p_1$, $p_2$, $p_3$ .... $p_k$ are given values (as given in the focus) of the variables $x_1$, $x_2$, $x_3$ ... $x_k$, respectively.

what comes naturally," are indeterminate, and, in any event, even if approximative limits may be set on the covert activities by obtaining a larger sample of relevant overt behavior, it is likely that such a person will show a "mongrel" strategy. That is, his covert behavior is likely to consist of redundant memory and conclusion drawing procedures such that he is capable of alternative routes to problem-solution, depending upon current conditions and unknown priorities that are a function of past experience. Thus, it is unlikely that clear discriminations may be made among persons who can perform the goal behavior without training. If, on the other hand, one attempts to discern what the superior problem-solver does by comparing those who can perform goal behavior with those who cannot, one is likely to learn that poor problem-solvers are forgetful, illogical, disorganized, or stereotyped in their behavior, in various orders and combinations depending upon the task, and that the superior problem-solver's success is a function of the contrasting virtues. Such finer discriminations as are made are more likely to obtain from the theory that is imposed than the observations themselves. In short, it is argued that the best way to study problem-solving behavior is to teach specified alternative strategies for solving a class of problems and then compare the resulting performances. If one is unprepared, at least, to specify alternative strategies capable of generating a certain performance, one is unprepared to observe problem-solving behavior as it occurs in nature intelligently anyhow.

One of the aims of this section is to formulate what children should learn if they are to acquire an inductive reasoning capability, more specifically, a capability for attaining conjunctive concepts within small closed systems. Perhaps the best strategy -- on strictly _a priori_ grounds -- is one that involves processing information available in both positive and negative instances, drawing conclusions on the status of variables, remembering the current conclusion after each trial instead of instances, and integrating instance selection and progress tracking with conclusion drawing. A diagram of such a strategy appears in Figure 9.

## 3. A PROCEDURE FOR TEACHING THE METHOD OF DIFFERENCE

Folklore has it that children are incapable of "abstract thinking." Indeed, the developmental literature does show that children do less well than adolescents and adults on most intellectual tasks. Piaget, the leading theorist in cognitive psychology, has been almost exclusively occupied with conceptualizing the development of intellectual functioning. His work permits the inference that it would be difficult for a child to learn involved kinds of problem-solving behavior.

It is also true that attempts to teach young children "advanced" forms of thinking generally have been unsuccessful, at least insofar as transfer of training has been the goal. For example, Ervin (1960) taught third and fourth graders a procedure for hitting a target on the rebound by aiming at a backstop. Most of the children acquired the behavior, but there was no overall difference between the groups that received training and a control group on a new task that required aiming a light at a mirror in such a way that the light would reflect on a target.

Despite the data that shows that children do relatively poorly on intellectual tasks and despite the failures of previous efforts to accelerate children's development of intellectual operations, there is reason to believe that it is possible to teach children important problem-solving skills. After all, almost all children learn a language.

Learning an inductive reasoning capability should not be inherently more difficult, if that which is to be taught is properly conceived. This assertion does not have a Utopian or revolutionary ring since there is currently a great deal of optimism about what children are capable of learning. The optimism was given its most dramatic expression by Bruner (1960, p. 33) in his now famous (or infamous) assertion "that any subject can be taught effectively in some intellectually honest form to any child at any stage of development." Many projects are now trying in one way or another to make a dramatic improvement in intellectual functioning. The ultimate outcome of these efforts cannot be forecast. There is room, however, for the suspicion that a number of the contemporary projects will come to grief.

In the first few decades of this century, repeated efforts were made to train "mental faculties." Courses in Latin and plane geometry, as well as specially designed "exercises," were found to have little generalized effect on intellectual functioning. The concept of "mental faculty" is not unlike the notions of many of those currently trying to improve basic intellectual functions. All of the emphasis is on an adaptive form or quality -- such as intelligence or creativity -- to the neglect of specific behavior in specific contexts. The writer is convinced that it is precisely those "basic capabilities" or "talents" that can be conceived as specific skills that offer the best odds for deliberate improvement. One modifies behavior rather than the forms of behavior.

To be sure, the term basic intellectual function implies a generalized capability available in a variety of contexts. How does one teach for generality? There is an answer to this question in the research on "learning to learn."

Harlow's (1949) fascinating work on learning sets showed cumulative nonspecific transfer of training as monkeys performed on a long series of tasks. One study involved a series of two-object discrimination problems. In each problem, one of two objects concealed a raisin. The naive monkey required a number of trials before consistently choosing the object concealing the raisin. After performing on a large number of problems, the monkey always chose the object concealing the raisin after just one trial. The monkeys had, in Harlow's words, "learned to learn." The monkeys had clearly not acquired a response to any specific stimulus, such as a certain size, shape, or color, for in any problem the positive object usually had as much in common with the negative objects of previous problems as it did with objects which previously had been positive. The common element running through all problems was a formal or schematic property. The schema came to provide the discriminative stimuli.

What exactly is learned when an organism learns to learn is a moot question. The dominant view is that the organism learns "observing responses" (Kurtz, 1955) or a "perceptual set" (Eckstrand and Wickens, 1954) though there are those who believe that "rules" are learned.

One point is clear. Stimulus or task variety during training is associated with generalized, positive transfer of training (Dashiell, 1924; Crafts, 1927; Wolfle, 1936; Callentine and Warren, 1955; Duncan, 1958; Morrisett and Hovland, 1959). By whatever mechanism, varied training modifies stimulus control. Behavior comes under the control of an abstract property (or abstract properties) of tasks.

Overview of training procedure. The general plan of instruction can now be sketched. First, the attempt was made to teach six-and seven-year olds terminal Method of Difference behavior with a roughly, sequential small-step teaching program. Second, the attempt was made to vanish the control of particular task characteristics and bring Method of Difference behavior under the control of certain abstract or formal task properties.

The program for the first phase of training consisted of seven large units or chapters that were called "games." Each game was designed to teach one or more of the component operations required for terminal Method of Difference behavior. The first three games were designed to set up appropriate conclusion-drawing behavior. The last four games were intended to teach instance selection behavior and to coordinate instance selection with conclusion drawing.

The second phase of instruction vanished the control of particular task cues on Method of Difference behavior. The hope was to reach a state in which Method of Difference behavior could be emitted for any task that was a small, well defined system and within which the problem was to attain a conjunctive concept. To this end, an abbreviated form of the program employed in the first phase of instruction was presented with each of five additional tasks that involved different variables, different numbers of variables, different arrangements of materials and apparatus, and different language -- that is different "story lines." The common element running through all tasks was a formal or schematic property.

The remainder of this section presents details of training procedure, describes the program, and outlines the training tasks.

Details of training procedure. The program was presented by human teachers. Programed instruction, with which this study is identified, usually involves a mechanical teaching agent. There is no reason in principle why human teachers could not present programs. Indeed, Skinner (1954) has likened a teaching program to what a skilled, private tutor does. There was a compelling reason for using human teachers rather than machines in the present case. There are no existing teaching machines with which the program could have been presented. In the first place, in the later stages of the program, teacher behavior is contingent upon sequences of student behavior; a computer-controlled teaching machine would be required to handle these contingencies. In the second place, there is no existing teaching machine that can present the stimulus materials and evaluate the constructed responses required by the program for students who can neither read nor write.

The attempt was made to have human teachers behave as though they were teaching machines, that is, the experimental teachers were given a program containing a specification of every detail of teaching behavior and were expected to follow the program slavishly. One advantage that teaching machines have as compared with human teachers was quickly discovered. The teachers could not decode the detail of the program and simultaneously attend to what the children were doing and saying. The pace of instruction was slow, in itself a liability when the subjects are first-graders who frequently have short attention spans. When the teachers were trying to follow the program slavishly, they were sometimes caught reading the next item instead of listening and watching. As a result, the teacher occasionally had to ask the child to repeat and sometimes behaved inappropriately. In view of these problems, the decision was made to abandon machine-like presentation of the program and take advantage of the capabilities of humans. For the same reasons, the teacher was not expected to keep a frame by frame record of subject behavior. The mode of presentation of the program that finally evolved and which was employed in the training of experimental subjects embodied the following rules:

1. The teacher followed in detail the general plan of the program -- that is the sequence and structure of the large units called games.

2. The teacher employed in detail the introductory frames for each game and any major transitional frames within games.
3. When a number of frames of similar type were to be presented, the teacher created the frames on the spot according to a rule of sequence. (Thus, subjects were presented with different frames and different sequences of frames). The likelihood that a subject will respond appropriately to a given frame is partly a function of the cues in the preceding frame. The teacher created and presented items such that item to item cueing was maximized at the beginning of the sequence and minimized at the end.
4. The teacher judged the point at which the subject had mastered whatever it was that a series of similar frames was supposed to teach. (Hence, the number of frames to which subjects responded was variable).
5. The teacher created frames on the spot according to a remedial rule, whenever a subject behaved inappropriately. Since there were several remedial rules, depending upon the section of the program, these will be detailed below.
6. The teacher followed in detail those items that were designated as criterion frames and also recorded subject behavior on criterion frames. The teaching that took place in this study may be generally characterized by saying that the behavior of the teacher was specified in considerably more detail than in the typical

study of teaching style (e.g., learner-centered vs. teacher-centered), but in less detail than in programed texts or teaching machine programs.

Three related card array tasks, designated Card Array I, Card Array II, and Card Array III, were employed in the first phase of training. Each of the tasks involved the same three variables; namely, color, number, and shape. In Card Array I, which is represented in Figure 1, each of the variables took on two values -- red or green, one or two figures, "box" or "diamond." Card Arrays II and III involved a third value of each variable. Figures were red, green, or blue. There were one, two, or three figures on each card. The figures were shaped like "boxes" (rectangles), "diamonds," or "circles" (ovals). Card Arrays II and III were identical except for the arrangement of cards. Array II was laid out in an orderly fashion,whereas III was arrayed in a disorderly manner. Each of the arrays was scotch-taped to a Masonite panel.

Game 1. Experimenter names concept. Subject points to all of the positive instances of the concept.

Instructions: "Would you like to play a game with these cards? I will name a secret. You point to all of the cards that show my secret. For example, if my secret were red, then you would point to these cards (E points to each of the red cards in Array I). Alright, my secret is green. Now you point to the cards that show my secret."

E presented concept names until, in his judgement, the S could point to the exemplars of every conjunctive concept that could be formed in Array I. At this point, E introduced Array II. When in E's judgement S could point to the exemplars of any conjunctive concept that could be formed in Array II, Array III was introduced and concept names were presented until the same criterion had been reached.

Remedial steps: If S (a) failed to point to all the cards that showed the secret, (b) failed to respond or, (c) pointed to the wrong cards, E presented remedial steps. If S failed to point to all of the cards that showed a secret, E said "Remember, you're supposed to point to all of the cards that show the secret." If S persisted, E said "My secret is ___________. You pointed to these cards (E points). Point to the other card(s) that show my secret." If S either failed to respond at all or responded incorrectly, E told S the answer to an analogous problem and then presented the problem upon which the error was made a second time. An analogous concept differs in the value of one relevant variable from the concept upon which an error was made and is otherwise the same. Example of procedure: S failed to respond to one diamond. E says "If my secret were one box, these cards (E points to appropriate cards) would show my secret. My secret is one diamond. Point to the cards that show my secret."

Game 2. Experimenter points to all of the exemplars of a conjunctive concept. Subject names concept.

Instructions: E congratulated S for his excellent progress, stating that S is doing so well that he is ready to play a harder game. "In the game we're going to play now, I will point at the cards and you name the secret the cards show. So, if I pointed at these cards (E points to 1Rb, 1Gb, 2Rb, 2Gb) then you would say box. Do you understand? Let's see if you can do it. These cards show the secret (E points to 1Rd, 1Gd, 2Rd, 2Gd). You say the secret. (S says diamond). Good!"

E presented problems until, in his judgement, the requisite level of mastery -- such that S could name the concept given a full set of exemplars no matter what the concept -- had been reached on Array I, Array II, and Array III.

Remedial steps: If S failed to respond or responded incorrectly to any Game 2 problem, E told S the answer to an analogous Game 1 problem and then presented the Game 2 problem upon which the error was made a second time.

Game 3. Experimenter points to each of a set of positive and negative instances that define a conjunctive concept. Subject names concept.

Instructions: E praised S for doing well and asked S if he thought he could play a harder game. Then E said, "In this game I will point to some cards. Some of the cards I point to will show my secret; some of the cards will *not* show the secret. You have to name the secret. For example, if I said this card (1Rb) shows my secret, this card (1Rd) does *not* show my secret, and these cards (2Rb, 1Gb) *do* show my secret, then you would say 'the secret is box.' Now you try it. This card (1Rd) shows the secret, this card (1Rb) does *not* show the secret, and these cards (2Rd, 1Gd) *do* show the secret. What is my secret? (Subject says *diamond*). Very good! That's right."[4]

*E* presented sets of defining instances from Array I until, in his judgement, *S* could respond with the appropriate name, no matter what the set of instances. Then *E* proceeded to the same criterion with Array II and Array III.

Remedial steps: If an *S* failed to respond or responded incorrectly, *E* presented an analogous problem, then presented the problem upon which an error was made again. An analogous problem in Game 3 involves a concept in which one value of one relevant variable is different from the concept upon which an error was made. The order in which the values are changed in the series of defining instances and the values of irrelevant variables are identical.

---

[4] A further explanation of the manner in which *E* composed sets of instances that defined conjunctive concepts is needed. The first instance was always positive and may be regarded as the focus instance. In the subsequent three instances, each value in the focus was changed once such that each of the three instances was the same as the focus with regard to two values and different with regard to one.

Game 4. Experimenter points to focus instance. Subject points to instances that differ in just one value from the focus instance.

Instructions: E says "Let's play a new game. These two cards (E points to 1Bd and 1Rd in Array II) show the same number and the same shape but they are different colors. Can you point to a card that shows the same number and shape as this card (2Rd) but is a different color? (S points to 2Bd or 2Gd). Good!"

E presented cards until he was satisfied that no matter what the focus instance, S could pick a card of a different color but otherwise the same. Number and shape were treated in identical fashion. Then E said "I'm going to make the game harder. When it is your turn, you have to point to a card that shows the same number and shape as my card and a different color, and another card that is the same color and shape as my card and a different number, and another card that is the same color and number but is a different shape." Problems were presented until, in E's judgement, no matter what the focus, S could pick three cards differing from the focus in number, color, and shape (any order), respectively and otherwise the same as the focus.

Remedial step: When S erred, E indicated the error and asked S to try again.

Game 5. Experimenter points to focus instance. Subject selects three instances that differ from the focus with regard to number, color, and shape, respectively, and are otherwise the same. Experimenter defines a conjunctive concept with the instances the subject has selected by assigning positive or negative to each of them. Subject names concept.

Procedure: E says "Let's make the game even harder. After you pick the cards I will say which ones show the secret and which ones don't. Then you tell me the secret." Focus instances and concepts are presented from Array II until S has attained mastery. Then problems are presented from Array III until S has attained mastery.

Remedial steps: Each Game 5 problem was presented in two parts. First, E presents the focus, and S selects a set of instances. If S fails to select a complete and proper set of instances, E describes the failure and asks S to try again. Second, when, and only when, a "legal" set of instances has been selected, E categorizes the instances as showing or not showing the concept. Then S attempts to name the concept. If S fails, an analogous task (see Remedial Step, Game 3) is presented and then S trys a second time to name the concept.

Game 6. Experimenter points to a focus instance. Subject selects three instances that differ from the focus with regard to number, color, and shape (any order), respectively, and are otherwise the same. After each properly chosen instance, the experimenter immediately indicates whether the instance "shows the secret" or not. Subject names the concept.

Procedure: E says "I'm going to change the game (Game 5) a little. Each time you pick a card, I will tell you whether it shows the secret right away." Otherwise, the procedure was the same as for Game 5.

Remedial steps: Remedial steps paralleled those of Game 5.

Game 7. Experimenter points to a focus instance and categorizes instances selected by the subject as showing or not showing the concept. Subject selects instances and names concept.

Procedure: E says "I think you're ready to play a new game. I will think of a secret and point to a card that shows my secret. You try to figure out my secret by picking cards. Each time you pick a card, I will tell you if it shows the secret. As soon as you are sure you know, tell me the secret." Arrays I, II and III were employed in Game 7. For each array in succession, E composed and presented problems until, in his judgement, no matter what the focus instance or what the concept, S could select three instances and state the conclusion such that the instances implied the stated conclusion and no other. When this level of mastery, in E's judgement, had been reached with Array III, S was presented with problems from the first phase training criterion list until he had attained the concept in three trials on seven out of eight consecutive problems. The criterion list was the same for all subjects and E kept a detailed record of S's performance.

Remedial steps: (1) If S selects an instance that differs in more than one value from the focus instance or other positive instance, call the instance negative. Note that application of this rule may involve changing the listed concept. (2) Call all consistent statements of the concept wrong. Application of this rule may also involve changing the listed concept.[5]

What has the child learned who has completed the first phase of training? All 30 of the children who received training in the study (also many other children who received training while the training procedures were being developed) reached criterion. That is, on seven out of eight consecutive card array problems each of the subjects attained conjunctive concepts in exactly three trials, such that the trials implied the stated concept and no other. The first answer to the question is that the terminal behavior was learned.

In the preceding section, it was argued that the best strategy for attaining conjunctive concepts within small, closed systems involves processing information available in both positive and negative instances, drawing conclusions on the basis of variables, remembering the current conclusion after each trial instead of instances, and integrating instance selection and progress tracking with conclusion drawing.

---

[5] The term "shaping" in the title on the project has its origin in these two rules. The first training procedure that was evolved moved quickly to a Game 7-like task with a very simple two-variable card array. Subjects' behavior was modified through procedures such as those involved in these two rules. Shaping is not a very prominent feature of the current training procedures.

Such a strategy, illustrated in Figure 9, is probably not a very good description of the behavior of most of the subjects who received training. Based on an analysis of the instructional program and inference from the observable behavior (in addition to that explicitly called terminal behavior), the strategy of the typical child who received training can probably best be described as involving processing of information available from only positive instances, memory for instances, conclusion drawing following complete selection of instances, conservative processing instance selection and progress tracking procedures. The argument that this strategy best describes the covert behavior of subjects is as follows. First, the strategy that was learned probably involved memory for instances and conclusion drawing after a complete set of instances for this procedure was heavily practiced early in training (Games 2, 3, and 5). In fact, no other procedure could have been employed until late in training (Game 6). Second, conclusion drawing was probably based on common elements of positive instances. Game 2 predisposed the subjects toward a common elements procedure, irregardless of entering behavior. Third, a conservative focusing instance selection and progress tracking algorithm was taught. No other procedure could yield successful performance in Game 4, though, behavior resulting from alternative procedures would have been reinforced in later games, if emitted.

There was actually considerable variability in the topography of subjects' behavior following the first phase of training, suggesting variability in strategies.[6] No doubt some of the variability is due to variability in the details of training. Presumably, if every subject was presented with exactly the same sequence of problems within the games that were described, there would have been greater topographical uniformity and, by inference, greater uniformity in strategies. A second source of variability is to be found in the training viewed more broadly. The games, the broad units of the program did not generally preclude alternative routes to the same terminal behavior. For example, in the preceding paragraph, it was asserted that the children probably employed a common elements procedure for drawing conclusions; however, not to be overlooked is the fact that these particular training procedures did not rule out any of the possible conclusion drawing algorithms.

An effort was made to teach a strategy bearing a closer resemblance to the one diagramed in Figure 9. A game was developed in which the experimenter presented three instances, indicating which of them showed the concept and which did not. The subject was expected to pick a fourth instance and then state the concept. The hope was that the subject would learn an instance selection procedure integrated with

---

[6] By "topography" of behavior is meant detailed features of overt, observable behavior not included in the definition of terminal behavior.

There was actually considerable variability in the topography of subjects' behavior following the first phase of training, suggesting variability in strategies.[6] No doubt some of the variability is due to variability in the details of training. Presumably, if every subject was presented with exactly the same sequence of problems within the games that were described, there would have been greater topographical uniformity and, by inference, greater uniformity in strategies. A second source of variability is to be found in the training viewed more broadly. The games, the broad units of the program did not generally preclude alternative routes to the same terminal behavior. For example, in the preceding paragraph, it was asserted that the children probably employed a common elements procedure for drawing conclusions; however, not to be overlooked is the fact that these particular training procedures did not rule out any of the possible conclusion drawing algorithms.

An effort was made to teach a strategy bearing a closer resemblance to the one diagramed in Figure 9. A game was developed in which the experimenter presented three instances, indicating which of them showed the concept and which did not. The subject was expected to pick a fourth instance and then state the concept. The hope was that the subject would learn an instance selection procedure integrated with conclusion drawing. The plan was to require the subject to pick two instances and then three after he was successful with one. The procedure did not work. None of the seven children with whom the procedure was tried were able to reach a criterion of mastery. Several other procedures were also tried, none of which worked. A completely new training procedure, designed to, among other things, teach an instance selection procedure integrated with conclusion drawing is currently being developed by the author. Based on past experience, this new procedure appears to have a good chance of success.

---

[6] By "topography" of behavior is meant detailed features of overt, observable behavior not included in the definition of terminal behavior.

*Vanishing the control of task characteristics*. Upon completion of the first phase of training, subjects received training on a series of five additional tasks. Problems composed from each of these tasks could be solved by the Method of Difference.

The five tasks were a collection of plastic cowboys, a collection of leaf facsimilies, a switching device, a ball-and-paddle device, and a pegboard game. These tasks are briefly described below:

*Cowboys*. Eight plastic cowboys, approximately two inches in height, are embedded in a plaster base. The cowboys vary in three respects; namely, horse vs. no horse, hat vs. no hat, and rifle vs. no rifle. A story-line was employed in the training. The problem is to determine "what kind of cowboy is a friend of the sheriff."

*Leaves*. Eight dittoed leaf facsimilies are attached with cellophane tape to 4-1/2" x 6" Masonite blocks. The leaves differ in three ways --

pinate vs. palmate, simple vs. compound, and notched edge vs. smooth edge. The problem is to determine "what kind of leaf shows the secret."

Switching device: A bank of six toggle switches and a master push-button switch face the subject on one side of a 8" x 6" x 5" metal cabinet. Any of the switches may be made relevant or irrelevant by manipulating switches at the rear of the cabinet. The problem is to "find which switches must be up in order for the buzzer to sound when the master button is pushed."

Ball-and-paddle device: A 14" x 7" x 3-1/2" wooden box has a channel through which a small ball may be rolled and six slots into which paint paddles may be inserted. When pushed in all the way, a solid paddle will stop the ball, whereas a notched paddle will permit the ball to pass. Notched and solid paddles may be placed in the slots in any combination. The problem is to find "which paddles must be out if the ball is to roll through."

Pegboard: A "railroad track," a "road" and a "pond" are painted in white on a 8" x 13" piece of green pegboard. Along the track, road, and pond are "trees" consisting of pegs with wooden balls glued to the tops. The trees vary with regard to location, height (tall or short), leaf color (white ball or red ball), and trunk color (yellow peg or green peg). The problem is to find the "kind of tree under which Pirate Pete buried his gold."

An abbreviated form of the first phase training was employed with each of the tasks that have just been described. The training sequence for each task consisted of an appropriately modified version of first phase Games 1, 2, 4, 6, and 7. Precisely the same instructional rules were employed as were used in the first phase of training. Tasks were presented in counterbalanced orders with random assignment of subjects to orders. When, in the E's judgement S had mastered a certain task, problems were presented from a criterion list of problems for that task until S reached a criterion of seven of eight consecutive solutions in exactly as many trials as there were variables in the task.

## 4. THE MEASUREMENT OF PROBLEM-SOLVING BEHAVIOR

One reason for the primitive state of knowledge about problem-solving is the lack of a satisfactory datum. Whether or not a problem is solved is frequently employed as a measure. When a large number of short problems, such as items in an aptitude test, can be presented, there is no doubt that number of solutions is a satisfactory measure. But when one or two long, complex problems are to be presented, simply noting success or failure is inadequate. First of all, no discrimination is made between the subject who solves the problem in a direct and efficient manner and the subject who stumbles upon the solution or who makes a lucky guess. Second, when only success or failure is noted, reliability is likely to be low (Anderson and Anderson, 1963). Third, it is difficult to find problems that are not either much too easy or much too difficult (Staats, 1957). If training is involved, a performance asymptote may prevent training effects from appearing.

Some of the difficulties of counting solutions can be circumvented by employing a subsidiary measure. Time to solution is a favorite, though other indices such as "number of hints required" (Gagné & Brown, 1961) are occasionally used. Such measures as time to solution do allow finer discriminations among subjects than counting successes and do avoid the asymptote problem. However, the fast subject is not necessarily efficient nor the slow subject necessarily confused and aimless. A subject may be slow because he is fumbling or because he is deliberating each move carefully (Duncan, 1963).

The investigator may make a global judgement of the adequacy or character of problem-solving based on an entire episode. The controversial clinical method of Piaget is an example of this approach. It does not really yield a measurement but rather an interpretation.

The investigators involved in computer simulation of problem-solving; notably Newell, Shaw, and Simon (1958), have resurrected introspection, long taboo in psychology, as a source of data. Thus far, introspective data is being used solely as a source of ideas for constructing the simulation, in other words, for building the theory. As Hovland (1960) noted, the task of assessing the degree of fit between the simulation and actual problem-solving protocols has not been faced. The usefulness of introspective data for this task remain to be seen, though on the basis of previous studies limited usefulness is to be expected.

The methodology of Bruner, Goodnow, and Austin (1956) will be considered in some detail, since it was a source of inspiration for the methodology employed in this study. They have argued first that one must consider the "sequence of responses if one is to appreciate the unfolding interplay between successive responses in reaction to prior consequences"(p. 243). Second, they have emphasized that it is necessary to "externalize the component steps or decisions in order to analyze behavior sequences" (p. 243). The concept attainment task developed by Bruner, Goodnow, and Austin, and adapted for the present investigation -- in which the experimenter presents a positive instance while the subject chooses instances and names the concept -- is admirably suited to "externalizing" sequences of problem-solving behavior. Moreover, this sort of concept attainment task is representative of a large class of important "real" problems.

The major theoretical innovation of Bruner and his associates was the concept of strategy. The measurement of strategies, however, left a great deal to be desired. Two varieties of "selection strategies" were hypothesized; namely, "scanning" and "focusing." The scanner alledgedly attempts to eliminate possible hypotheses. He chooses instances to eliminate the largest possible number of hypotheses. Upon learning whether an instance is positive or negative, the scanner must compute which hypotheses have been eliminated and which are still possible. The focuser, on the other hand, allegedly tests the relevance of variables. The strategy is called focusing because the ideal focuser is said to vary just one value of the initial positive instance with each instance he chooses. In practice, it was asserted that a modified form of focusing was employed in which the subject changed one and only one value of the focus instance or other positive instance at each choice (p. 94).

How did Bruner and his associates decide whether a subject was a focuser or a scanner? Their procedure was described in a single footnote as follows, and never further elaborated. "A focuser was defined as a subject whose choices in the main varied only one attribute value from those attribute values of the focus card that had been found relevant or were still untested. The only other types of choices permitted subjects in this group were redundant choices where the third value of an attribute was checked, or focus-gamble choices to be utilized as such. When the majority of the choices was of these types, the subject was considered a focuser. All others were treated as scanners (note p. 95)."

The first question one might ask about such a procedure is to what extent did independent raters agree in their classification of subjects. This question was never answered, and, because it was not, the implication is that anyone who looked at the protocols would certainly classify subjects the same way. Experience has shown that the strategies employed by subjects are not self-evident to raters. Furthermore, first graders, at least, seem to fall along a continuum with respect to the choices of instances they make. Forcing the data into two classes seems arbitrary and seems to beg the theoretical question of the kind of strategy involved, or indeed, whether the concept of strategy is viable at all.

Is there a way to assess complex problem solving episodes in a manner that adequately characterizes sequences of behavior but at the same time avoids the vagaries of global, subjective judgement or arbitrary simplification. To this end, a methodology was developed that includes standardized presentation of problems in which much of the subject's behavior and all of the allowable experimenter behavior is recorded in coded form and content analysis of problem-solving protocols that emphasizes low-inference categories. The rest of this section details the presentation procedure, describes the content analysis system, and presents reliability coefficients and other data that bear upon the stability of the measures.

Presentation of problems. A procedure was developed for presenting problems that proved suitable -- with minor modifications -- for all of the various problems that were employed in the investigation. In every case, the subject was first oriented to the task. Orientation consisted of the presentation of 20 conjunctive concepts that could be formed within the task. The experimenter said the name of each concept; the child was expected to point to each of the positive instances. If the child made an error on any concept, the experimenter pointed to the positive exemplars of an "analogous" (similar) concept, then presented the concept upon which the error was made again. Also, any concept upon which an error was made was presented again at the end of the list.

Orientation was expected to serve several functions. Chief among the functions was a sort of definition of conjunctive concept. Several attempts to give a criterial definition of conjunctive concept in simple language failed. Definition by example seemed to work fairly well.

At least when this procedure was employed, most of the concepts children stated, whether correct or incorrect, were conjunctive. Conjunctive concepts only were stated on 722 of the 780 problems that were presented.

After the orientation in which the child learned -- or demonstrated knowledge of -- the conjunctive concepts that could be formed within the task, two problems from the task were presented. In every case, the problem was to attain a conjunctive concept. Both of the concepts to be attained in problems had appeared twice in the list of concepts presented during orientation. The experimenter always began the problem by presenting a positive instance. After that the subject selected instances until he could name the concept.

Instructions. Instructions varied somewhat depending upon the nature of the task. There is a brief manual that accompanies each task that contains verbatum instructions, coding of variables, and procedural detail (Screws and bolts problem, Anderson, 1963; Housing development problem, Anderson, 1963; Pendulum problem, Vojack, 1963; Chemical problem, Fraunfelker, 1963; Pencil problem, Anderson, 1963; Cowboy problem, Anderson, 1963; Pegboard problem, Anderson, 1963). The instructions employed with Suburbia will be set forth for illustration. After orientation, the experimenter said "Now we're going to play a new game. I will think of a secret, one of the secrets we had in the first game. Your job is to figure out the secret. Here's how you do it. I point to one house that shows my secret. You pick other houses.

Each time you pick a house, I will tell you whether or not it shows the secret. As soon as you are sure you know the secret, tell me what it is. Don't pick any more houses than you really need to to figure out my secret." Instructions were virtually identical for the other tasks. The seven tasks and the problems involving these tasks are briefly described below.

Screws. Materials consist of 16 screws and bolts that differ in terms of four attributes as follows: short (1-1/2") or long (3"); round-headed or flat-headed; steel or copper; screw or bolt. The materials are placed in a haphazard pile on a table in front of the subject. The experimenter mixes the screws and bolts after the first problem. Each problem involves figuring out which kind of screws and bolts "shows the secret." The concept to be attained in the first problem is flat head. The focus instance is a short, flat-headed, copper screw. The concept to be attained in the second problem is screw. The focus instance is a long, round-headed, steel screw.

Suburbia. Materials consist of sixteen miniature houses and other fixtures glued to a 16" x 16" Masonite panel in a disorderly arrangement. The houses differ with regard to four variables: one-story vs. two-story; north side of the road vs. south side of the road; tree vs. no-tree; car vs. no-car. Problems consist of determining which kind of house "shows the secret." The concept to be attained in the first problem is South side of the street with tree and no car. The focus instance is a one-story house on the south side of the street with a tree and no car. The concept to be attained in the second problem is two-story house with a tree and a car. The focus instance in the second problem is a two-story house on the North side of the street with a tree and a car.

Pendulum. Materials consist of a metal stand 40 inches high, capable of supporting three pendulums simultaneously, three pendulums of different lengths such that the period of the shortest is .5 seconds, the period of the one of middle length is 1 second, and the period of the longest is 1.5 seconds. Included also are two balls of 1" diameter, one made of steel and one of cork. Attached to the back of the stand is a piece of postal board 22" x 28" on which are drawn two lines of 30° and 45° from the perpendicular. There are three variables, the length of the pendulum, the amount of weight attached to the pendulum, and the angle from which the pendulum is released. The problem is "to figure out what makes this swing the most times in 10 seconds." The focus instance is the demonstration that the heavy weight on the middle length string released from a 45° angle swings 10 times in 10 seconds.

Chemical. Materials consist of a four-ounce amber bottle, a watch glass, and a set of five one-ounce dropping bottles labeled A, B, C, D, and E. Two of the latter bottles contain water. The other three contain, respectively, dilute solutions of sulphuric acid, hydrogen peroxide, and potassium iodide. The problem is to find "which bottles you have to use to get yellow." The concept to be attained in the first problem is that bottles B, C, and D are necessary. The concept in the second problem is B, C, and E. In each case, the focus instance is the demonstration that if drops from each of the five bottles is placed in the amber jar and then the contents are poured on the watch glass, the liquid is yellow.

Pencils. Materials consist of 12 yellow, hexagonal, No. 2 pencils. The pencils display three attributes: length (6 inches, 4 inches, 2 inches); presence or absence of eraser; sharpened or unsharpened. The pencils are arrayed in a disorderly fashion on a table in front of the subject. The problem is to figure out which kind of pencil "shows the secret." The concept to be attained in the first problem is unsharpened without eraser. The focus instance is a long, unsharpened pencil without an eraser. The concept to be attained in the second problem is medium length with eraser. The focus instance in the second problem is a medium length, sharpened pencil with an eraser.

Cowboys. The cowboys task was described in the preceding section since it was employed as a training task. Problems involved figuring out which kind of cowboy is "a friend of the sheriff." The concept to be attained in the first problem is no rifle and hat. The focus is a cowboy on a horse with a hat but no rifle. The second problem concept is standing with rifle. The focus instance is a cowboy standing who has a rifle but no hat.

Pegboard. The pegboard task was also employed during training and is described in the preceding section. Problems consist of determining "the kind of tree under which Pirate Pete buried his gold." The concept to be attained in the first problem is tall with red leaves and green trunk. The focus instance is a tall tree by the road that has red leaves and a green trunk. The second problem concept is short by the pond with yellow trunk. The focus instance for the second problem is a short tree by the pond with white leaves and a yellow trunk.

Experimenter behavior during the presentation of a problem was prescribed. The following three and one-half pages are quoted verbatum from the Manual for presenting logical problems (Anderson, 1963).

> During the problem. The experimenter should maintain a pleasant but neutral role while the child works on the problem, If necessary, remind the child of procedures (e.g., wash out the bottle), otherwise say nothing except as indicated below.
>
> A. The child stops short of solving the problem.
>
> It is the experimenter's job to keep the child performing, if he can, until the child has either solved the problem or exhausted the maximum number of trials allowed. Short of these criterion, if the child stops trying to solve the problem and it is apparent he is not going to continue on his own, wait 10-15 seconds. Then if he still isn't doing anything, give a slight motivational prompt (M 1): What are you going to do now?
>
> Wait 5-10 seconds. If the subject still is not performing give a moderate motivational prompt (M 2): See if you can figure out .... The ending of M 2 depends upon the problem, Some of the suitable endings are as follows:
>
> ... my secret by pointing at more pencils,
> ... which bottles you have to use to get yellow by mixing chemicals and pouring them on the glass,
> ... which switches must be up to make the buzzer work; try more switches.

Wait 5-10 seconds. If the child still is not performing, give a strong motivational prompt (M 3): Exhort the subject to continue. The wording of the exhortation is up to the experimenter. In general encourage the subject and reiterate the method he is to use. Illustrative M3: "Come on. You can do it. Just mix the chemicals in this brown jar and then pour what you have in the jar onto the glass. When you know which bottles you have to use to get yellow, tell me what they are. I'm sure you can figure it out. Let's see you try." Exhort the subject until he performs or until it is obvious he's not going to try no matter what you say.

B. The subject states an invalid conclusion and acts as though the problem is solved.

If a subject states an invalid conclusion, whether the conclusion is consistent or inconsistent with the evidence at hand, wait for 10-15 seconds.

If the subject continues to perform, say nothing. However, if he acts as though he has solved the problem (e.g., does nothing, sits back in his chair, looks expectantly at you) then give a slight challenge prompt (P 1): Are you sure ...? The ending of P 1 depends upon the problem and upon the particular invalid conclusion the subject has given. The following are examples of P 1 that would be appropriate under certain circumstances: Are you sure that the secret is long-pencils-with-erasers? Are you sure that A, B, C, and D have to be used to get yellow? Are you sure that switches #1, #5, and #6 have to be up to make the buzzer work? In short, you are asking the child if the evidence compels him to accept the conclusion he has stated. The use of the imperative form might be confusing with collections or arrays; with these problems communicate the imperative idea by the inflection in your voice.

If the child agrees that he is not sure, wait 10-15 seconds. If he begins to perform, say nothing more. If he fails to continue give M 1.

If, after P 1, the child insists he is sure or if he says or does nothing, give a moderate challenge prompt (P 2): Maybe .... The ending of P 2 depends upon the problem and the particular conclusion that was stated. Following are examples that would be appropriate under some circumstances: Maybe silver screws with round heads is not the secret. Maybe the period is not changed by the length of the string and the amount of weight. Maybe Pirate Pete did not bury his gold under trees with red leaves by the pond.

If the subject agrees "maybe not," but does not continue after 10-15 seconds, give M 1. If he begins to perform say nothing more.

If, after P 2, the subject insists his conclusion is correct or if he says and does nothing, give a strong challenge prompt (P 3): .... *are not* ..; *try to find out* ..... The rest of P 3 depends upon the problem and the particular conclusion that was stated. Samples: Short pencils with sharp points *is not* the secret; *try to find out* what the secret really is. A, B, C, and D are not the bottles that *have to be used* to make yellow; try to find out what bottles you have to use.

There is a special case of P 3. Sometimes through luck or contact with someone who has already completed the problem, a subject will state the "right answer" even though he hasn't proved it. In these cases P 3 should be worded as in the following example: Maybe B, C, and D are not the bottles you have to use to make yellow; try to find out for *sure* which bottles are needed.

C. The child states a bizarre conclusion.

Any conclusion of the child that is *not* a conjunction of the values of the variables of the problem on which he is working is a *bizarre conclusion*. Whenever a child gives a bizarre conclusion, the experimenter should give a modified *strong challenge prompt* (P 3) as in the following example: "No that is not my secret. We don't have any secrets like that in this game. The secret is one of the secrets we had in the first game."

D. Eliciting conclusions from subject.

When the child has accumulated enough evidence to draw a valid conclusion but does not volunteer one, the experimenter should give a slight eliciting prompt ($E_1$): As soon as you know ... The ending of the eliciting prompt depends upon the problem. Examples of endings follow:

.... my secret tells me what it is,
.... which bottles you have to use to make yellow tells me which ones,
.... which switches have to be up tell me.

Note that $E_1$ is given only once. If the subject fails to give a conclusion or gives an erroneous conclusion the session must continue.

If a child who has completed the maximum number of trials allowed has accumulated enough evidence to draw a valid conclusion but does not volunteer one, give a *strong eliciting prompt* ($E_2$): *I can't let you (select any more instances). Can't you tell me (the secret) now?* Of course the bracketed phrases change according to the problem.

*Terminating performance on a problem.* There are three conditions under which performance on a problem can be terminated as follows:

A. The child *refuses to continue* to try to solve the problem. He cannot be budged from his active or passive resistance by experimenter exhortation.

B. The child *completes the maximum number of trials allowed* without solving the problem. The maximum number of trials allowable is equal to the sum of the values across all variables. For illustration the pencil problem involves three different lengths, eraser or no-eraser, and sharpened or unsharpened. Thus a total of seven trials is allowed.[7]

---

[7] For obvious reasons, a subject cannot be allowed to continue indefinitely. A fixed limit on trials would not allow for differences in task difficulty. It has been shown (e.g., Archer, 1962) that the difficulty of a concept identification task is roughly an increasing function of the bits of information involved in the task. Therefore, the formula described above should set limits that vary approximately according to task difficulty.

C. The child *solves the problem and states a valid final conclusion* or the child solves the problem and states valid "part" conclusions for each variable. Solving the problem does not mean stating the "right answer." It means demonstrating that a certain conclusion and no other is true. In order to solve a problem a child must test the relevance of each variable and draw the proper inference.

When in doubt continue the session. If you are not sure whether a child has solved the problem keep the child performing. If the child really did solve the problem, the extraneous portion of the protocol can be eliminated by the person who does the scoring.

*Recording*. Everything the child says and does and everything the experimenter says and does must be recorded in order of occurrence.

A. Remarks that do not pertain *directly* to performance do not need to be recorded (e.g., "I think I've got it now," "Nothing I try seems to work.").

B. Trials are to be recorded using code symbols. Trials are to be numbered consecutively along the left margin.

C. Statements by the child regarding relevance and irrelevance of variables must be written out so that they will not be confused with trials. Confusion is a special danger in the chemical problem. Be sure to preface statement with "Need ...." or "Don't need ...." (i.e., "Don't need A or E.").

## Content Analysis of Problem Solving Protocols

In the classical concept formation task, the experimenter, *not* the subject, has control over the sequence of instances. With the procedure developed by Bruner, et al., (1956) and used in the present investigation, the experimenter presents the first instance and the subject selects every instance thereafter. Thus, the *sequence* of instances as well as the *number* of instances or trials is a datum. No doubt the sequence of instances is, in part at least, a function of the subject's strategy. The difficulty is that it is impossible to "see" a subject's strategy. In fact, as has already been shown, there are an indefinite number of strategies that can yield any given sequence of problem-solving behavior. Nonetheless, at a more molar level, there is a way of categorizing a particular trial that depends upon the sequence of previous trials.

A trial may be either *inefficient*, *redundant*, or *repetitious*. An inefficient trial yields less than the optimum amount of information about the concept. By either an information analysis or a logical analysis, the maximum amount of information is obtained from an instance that differs with regard to just one value from a previous positive instance. Any instance that differs in more than one way from each of the previous positive instances is *inefficient*. The truth of this statement does not depend upon the strategy which the subject is using. If a trial is inefficient, it is inefficient whether the subject is processing possible concepts, irrelevant variables, or necessary values.

A *redundant* trial yields only information available from previous instances. A trial is redundant if it yields a conclusion that is implied by the previous instances. Operationally, a trial is redundant (1) if the instance does *not* include all values previously shown to be necessary or (2) if the instance contains an as yet untried value of a variable known to be irrelevant and the instance is otherwise the same as a previous instance. Redundancy can also be defined in terms of information theory. As in the case of inefficiency, the truth of the assertions about redundancy does not depend upon the nature of the strategy being employed by the subject.

A repetitious trial exactly duplicates a previous trial.

Some redundant trials are also inefficient. The distinction between the two categories is maintained by the arbitrary convention that a trial is categorized as inefficient only if it is not redundant.

With these definitions and conventions, it is possible to classify instance selection behavior into four mutually exclusive and exhaustive categories. A trial is either inefficient, or redundant, or repetitious, or it is of high utility and is neither inefficient, redundant, nor repetitious.

The analysis of "concept naming" or "conclusion stating" behavior depends upon the analysis of instance selection behavior. A statement of the concept is either valid, consistent, inconsistent, or "bizarre." These categories are exhaustive of the unambiguous concept-naming behavior of subjects. Only unqualified declarative statements about the relevance or irrelevance of variables or unqualified conclusions about the concept are categorized. Any nonconjunctive statement is categorized as bizarre. Otherwise, statements are categorized, as the category names suggest, according to whether they are consistent or inconsistent with the instances that have been chosen up to that point. Note that a statement that is consistent at one point in time might be inconsistent at another and that a statement that is valid might be consistent at an earlier point. Valid statements about variables or values of variables are not categorized.

When a subject makes a valid statement of the concept, he has by definition solved the problem. In order to be counted as solving the problem, the subject has to select a series of instances such that a statement of the instances implies the concept he states and no other.

The entire content analysis scheme is summarized in Figure 12. Included in the scheme, in addition to the categories that have been discussed, are all of the kinds of contingent experimenter behavior and a few other categories that will not be discussed since, as it has turned out, they do not figure in much of the discussion that follows.

## Three Summary Measures of Problem Solving Performance

As will be seen shortly, there is no question but what raters can make the distinctions involved in the content analysis scheme that has been outlined. The utility of these fine distinctions is another matter. There might be some situations in which this degree of detail is desirable. Perhaps there will be a time when a sophisticated theory makes one prediction about number of redundant trials and a different prediction

Figure 12

Summary of Content Analysis Schema

| Category | Symbol |
|---|---|
| Subject performs logically inefficient trial | T1 |
| Subject performs logically redundant trial | T2 |
| Subject repeats trial | T3 |
| Subject makes consistent but invalid statement | SC |
| Subject makes inconsistent statement | SI |
| Subject makes bizarre statement | SB |
| Experimenter gives slight motivational prompt | M1 |
| Experimenter gives moderate motivational prompt | M2 |
| Experimenter gives strong motivational prompt | M3 |
| Experimenter gives slight challenge prompt | P1 |
| Experimenter gives moderate challenge prompt | P2 |
| Experimenter gives strong challenge prompt | P3 |
| Experimenter gives slight eliciting prompt | E1 |
| Experimenter gives strong eliciting prompt | E2 |
| Longest run of efficient trials | LR |
| Subject solves problem | S |
| Subject refuses to continue | RC |
| Total trials subject uses | TT |

about number of inefficient trials. For present purposes, it seemed desirable to sacrifice some degree of detail for stability and reliability. For this reason, three summary measures were chosen. Number of unnecessary trials is the sum of inefficient, redundant, and repetitious trials. Number of invalid statements is the sum of consistent, inconsistent, and bizarre statements. Number of solutions is the sum of number of problems solved.

An abortative effort was made to derive a single, global measure of good problem solving by summing the content analysis categories for each subject after weighting the categories in some way. The single-measure plan was eventually abandoned because it was too complicated, seemed to introduce artifacts into the data, and obliterated distinctions of possible utility for which there was a clear basis. As part of the program to obtain empirical weights for the content analysis categories, six factor analyses were performed (Walsh, 1963). Two card-array problems, two chemical problems, and two problems involving a switching device were administered to 54 third graders from a suburban school district in New Jersey. Problem-solving protocols were content analysed on the basis of an early version of the system previously described. Category frequencies were intercorrelated by problem. The correlation matrices were factor

---

[8] If a child refuses to continue, number of unnecessary trials is total allowable trials minus number of trials that are neither inefficient, redundant nor repetitious.

analysed by the principal components method. Unity was employed in the diagonals. From three to five factors with latent roots greater than 1.0 were obtained per problem. These factors accounted for from 67 per cent to 78 per cent of the cumulative trace. Additional calculations -- that will not be described, but which included graphical rotations to an *a priori* criterion -- were performed in order to obtain a set of weights. The point of describing the factor analyses is that they provide justification for the three summary, problem-solving measures. In five of the six analyses, the first three factors were similar and also quite clear and interpretable. At the time, they were called Successful Strategy, Length of Protocol, and Amount of Verbalization. These correspond to number of solutions, number of unnecessary trials, and number of invalid conclusions, respectively.

The categories that comprise the three measures of problem solving are anchored unambiguously in the logical and philosophical foundation of the study. In this sense, their validity is incontrovertible. Nor is the compilation into three measures entirely gratuitous. The factor analyses suggests a degree of empirical validity for these summary constructs.

## Reliability of Problem-Solving Measures

*Interrater reliability*. The first question about reliability that should be asked is the degree to which different raters agree in the scoring of problem-solving protocols. Table 1 reports interrater reliability coefficients for the three measures of problem solving for each of 13 problems presented to 60 first grade children. The protocols, identified only by number, were scored blind by two raters who worked independently. The coefficients range from .74 to .97 with a median of .90 for number of invalid conclusions. The range is from .86 to .99 and the median is .97 for number of unnecessary trials. Problem solution was scored with a reliability that ranged from .74 to 1.00 and had a median of .91.

If there is any reason for doing so, the narrow categories within the content analysis scheme can also be reliably scored. Two sets of coefficients were computed for illustration. Correlations between number of inefficient trials scored by two raters for the 13 problems ranged from .66 to .94 with a median of .82. The coefficients for number of redundant trials ranged from .63 to 1.00 and had a median of .92.

Shapes of the distributions of measures. Number of solutions was certainly not normally distributed, in fact, showed no uniform distribution of any kind by task. This fact should be kept in mind when interpreting the correlation coefficients that appear later in this section. When number of solutions to as many as five tasks are combined, the resulting distribtuion appears to be normal.

Tables 2 and 3 contain measures of kurtosis and skewness by task and treatment group for number of unnecessary trials and number of

### Table 1

### Interrater Reliability Coefficients for Three Summary Measures of Problem-Solving

| Problem | Invalid conclusions[a] | Unnecessary trials[a] | Solution[b] |
|---|---|---|---|
| Screws 1 | .75 | .97 | .91 |
| Screws 2 | .91 | .93 | .87 |
| Suburbia 1 | .96 | .97 | .89 |
| Suburbia 2 | .93 | .98 | .90 |
| Pendulum | .78 | .86 | .88 |
| Chemical 1 | .84 | .90 | .74 |
| Chemical 2 | .86 | .90 | .92 |
| Pencil 1 | .90 | .99 | .91 |
| Pencil 2 | .74 | .97 | .93 |
| Cowboy 1 | .91 | .93 | .94 |
| Cowboy 2 | .94 | .98 | .97 |
| Pegboard 1 | .97 | .99 | 1.00 |
| Pegboard 2 | .92 | .99 | 1.00 |

[a] Pearson product-moment correlation.

[b] Phi coefficient.

invalid conclusions, respectively. Unnecessary trials falls within normal limits rather consistently. Invalid conclusions shows a Poisson or near-Poisson distribution as is evidenced by the significant right skew seen in Table 3 for most cases. Table 4 contains measures of kurtosis and skewness for the square root of number of invalid conclusions. As can be seen, the distributions of transformed scores are, for the most part, more nearly normal.

Since there are some distributions that depart from normaltiy, some heterogeneity of form and occasional heterogeneity of variance, the one per cent level of confidence was chosen for all of the parametric tests to be reported. It is known that analysis of variance is not markedly affected when assumptions are violated. Hence, selection of one per cent as the apparent level of confidence should at least place the actual level of chance rejection of the null hypothesis within the customary five per cent limit.

*Internal-consistency reliability*. There seldom has been much concern with reliability on the part of investigators in the area of problem solving. Perhaps this is one reason progress has been slow. Usually there is no attempt to assess reliability with the exception, occasionally, of degree of interrater agreement. When correlations among measures of allegedly the same thing are obtained, the coefficients are frequently low. For instance, Anderson and Anderson (1963) obtained an average correlation of .20 between number of solutions to three problems of the sort developed by Maier and Duncker.

### Table 2

### Kurtosis and Skewness of Distributions of Number of Unnecessary Trials

| | Training Group<br>N = 30 | | Control Group<br>N = 30 | |
|---|---|---|---|---|
| Task | Kurtosis[a] | Skewness[b] | Kurtosis[a] | Skewness[b] |
| Screws | .85 | .10 | .71 | .72 |
| Suburbia | .84 | .29 | .76 | -.84 |
| Pendulum | .79 | -.29 | .82 | -.15 |
| Chemical | .80 | .01 | .78 | .92 |
| Pencils | .82 | .28 | .79 | -.79 |
| Cowboys | .86 | .57 | .64 | -1.61 |
| Pegboard | .81 | -.62 | .80 | -.12 |

[a] The measure of kurtosis is $a$ = (mean deviation)/(standard deviation). The one per cent confidence limits are $.88 > a > .71$.

[b] The measure of skewness is $\sqrt{b_1}$ for which the one per cent confidence limits are ± .98.

## Table 3

## Kurtosis and Skewness of Distributions of Invalid Conclusions

| | Training Group N = 30 | | Control Group N = 30 | |
|---|---|---|---|---|
| Task | Kurtosis[a] | Skewness[b] | Kurtosis[a] | Skewness[b] |
| Screws | .66 | 2.08 | .77 | .92 |
| Suburbia | .84 | .54 | .76 | 1.58 |
| Pendulum | .65 | 1.45 | .58 | -.02 |
| Chemical | .76 | .75 | .64 | 1.76 |
| Pencil | .81 | .21 | .80 | .66 |
| Cowboys | .74 | 1.77 | .70 | 2.05 |
| Pegboard | .77 | 1.21 | .76 | 1.49 |

[a] The measure of kurtosis is $\underline{a}$ = (mean deviation)/(standard deviation). The one per cent confidence limits are $.88 > a > .71$.

[b] The measure of skewness is $\sqrt{b_1}$ for which the one per cent confidence limits are ± .98.

Table 4

Kurtosis and Skewness of Distributions of the Square Root of Invalid Conclusions

| | Training Group<br>N = 30 | | Control Group<br>N = 30 | |
|---|---|---|---|---|
| Task | Kurtosis[a] | Skewness[b] | Kurtosis[a] | Skewness[b] |
| Screws | .75 | .04 | .82 | -.24 |
| Suburbia | .76 | -.52 | .85 | .12 |
| Pendulum | .78 | .05 | .69 | -1.30 |
| Chemical | .76 | -.78 | .63 | 1.04 |
| Pencil | .77 | -.72 | .87 | -.28 |
| Cowboys | .91 | .60 | .88 | .68 |
| Pegboard | .76 | -.27 | .91 | .58 |

[a] The measure of kurtosis is $\underline{a}$ = (mean deviation)/(standard deviation). The one per cent confidence limits are $.88 > a > .71$.

[b] The measure of skewness is $\sqrt{b_1}$ for which the one per cent confidence limits are ± .98.

If two highly similar problems involving the same instructions and the same task are presented one after another, surely, for any trustworthy index of problem-solving performance, there will be at least a moderate, positive correlation between the first problem measure and the second problem measure. Herein lies a minimum reliability condition. Table 5 contains first problem-second problem correlations for each of the three measures by task. The table presents the mean of the correlations obtained for the control group and the experimental group. Correlations computed from pooled data were inflated because of the treatment effect.

All subjects performed first on the cowboy problems, then on the pegboard problems, then on the remaining five tasks in counterbalanced orders. The two problems from a single task were always presented in the same order. The problems within a task were similar in the respect that each of them involved the same number of relevant variables. Across tasks the number of relevant variables per problem ranged from one to three.

The intratask correlations shown in Table 5 ranges from .04 to .67 with a median of .19 for invalid conclusions. The range is .26 to .41 and the median is .33 for unnecessary trials. The corresponding figures for solutions are .17 to .39 and .36. Hereafter, the first problem and second problem scores have always been combined. Median total scale reliabilities by task, estimated with the Spearman-Brown formula, are .32, .50, and .53 for invalid conclusions, unnecessary trials, and solutions, respectively.

Table 5

Correlations Between the Same Measure Scored from Two Similar Problems Involving the Same Task[a]

| Tasks | Measures | | |
|---|---|---|---|
| | Invalid Conclusions | Unnecessary Trials | Solutions |
| 1. Screws | .19 | .28 | .35 |
| 2. Suburbia | .51 | .28 | .37 |
| 3. Pendulum[b] | --- | --- | --- |
| 4. Chemical | .04 | .26 | .17 |
| 5. Pencil | .13 | .41 | .39 |
| 6. Cowboys | .67 | .38 | .29 |
| 7. Pegboard | .19 | .39 | .38 |

[a] The coefficients presented are the means (via z transformation) of experimental group (N = 30) and control group (N = 30) correlations. Correlations computed from pooled data are higher, but involve a spurious treatment effect.

[b] Only one pendulum problem was given.

Other relationships. Tables 6, 7, and 8 present correlations among tasks for invalid conclusions, unnecessary trials, and solutions. Each of the three tables contain first-order partial correlations. Experimental group-control group membership was partialled out to eliminate the treatment effect. Zero-order intercorrelations were considerably higher.

Table 9 presents correlations among measures within tasks. The focus here is upon the internal economy of problem-solving performance. Once again, experimental treatment was partialled out in order to avoid an inflated estimate of the relationships.

Five months before the presentation of problems, the California Test of Mental Maturity, Primary Form, 1957 Edition, was administered to all subjects. The test battery was administered and scored by the project staff rather than by classroom teachers. Table 10 contains the correlations of mental age with each of the measures on each of the tasks. Once again, experimental treatment has been partialled out. Parenthetically, correlations with chronological age were also computed, but were near zero in virtually every case.

There are several conclusions that emerge from the data that has been presented. First, there is little relationship between performance on the pendulum task or the chemical task and performance on the other tasks. Number of invalid conclusions on the pendulum task showed a mean correlation of .05 with number of invalid conclusions on the remaining

Table 6

Correlations Among Numbers of Invalid Conclusions by Task[a]

| Tasks | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1. Screws | --- | | | | | | |
| 2. Suburbia | .37 | --- | | | | | |
| 3. Pendulum | .03 | -.04 | --- | | | | |
| 4. Chemical | .07 | .13 | .03 | --- | | | |
| 5. Pencils | .15 | .23 | .12 | -.01 | --- | | |
| 6. Cowboys | .10 | .19 | .13 | -.11 | .33 | --- | |
| 7. Pegboard | .16 | .32 | -.01 | .16 | .38 | .24 | --- |

Note. -- Experimental treatment has been partialled out.

[a] Square root of number of invalid conclusions.

Table 7

Correlations Among Numbers of Unnecessary Trials by Task

| Tasks | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1. Screws | --- | | | | | | |
| 2. Suburbia | .12 | --- | | | | | |
| 3. Pendulum | .17 | .11 | --- | | | | |
| 4. Chemical | .21 | -.01 | -.12 | --- | | | |
| 5. Pencils | .28 | .58 | .14 | .16 | --- | | |
| 6. Cowboys | .16 | .23 | .31 | .10 | .32 | --- | |
| 7. Pegboard | .34 | .42 | .30 | .05 | .30 | .32 | --- |

Note. -- Experimental treatment has been partialled out.

Table 8

Correlations Among Numbers of Solutions by Task

| Tasks | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1. Screws | --- | | | | | | |
| 2. Suburbia | .23 | --- | | | | | |
| 3. Pendulum | -.10 | .06 | --- | | | | |
| 4. Chemical | .06 | .13 | -.02 | --- | | | |
| 5. Pencils | .26 | .33 | -.07 | .07 | --- | | |
| 6. Cowboys | .15 | .22 | .00 | -.05 | .27 | --- | |
| 7. Pegboard | .15 | .45 | -.03 | .30 | .15 | .34 | --- |

Note. -- Experimental treatment has been partialled out.

Table 9

Correlations Among Measures of Problem Solving Within Task

| | Tasks | Invalid Conclusions[a] vs. Unnecessary Trials | Invalid Conclusions[a] vs. Solutions | Unnecessary Trials vs. Solutions |
|---|---|---|---|---|
| 1. | Screws | .21 | -.26 | -.45 |
| 2. | Suburbia | .03 | -.13 | -.79 |
| 3. | Pendulum | .17 | -.60 | -.44 |
| 4. | Chemical | -.21 | -.34 | -.29 |
| 5. | Pencils | .19 | -.33 | -.58 |
| 6. | Cowboys | .47 | -.49 | -.79 |
| 7. | Pegboard | .32 | -.30 | -.80 |

Note. -- Experimental treatment has been partialled out.

[a] Square root of number of invalid conclusions.

Table 10

Correlations of Mental Age with Measures of Problem Solving by Task

| | Tasks | Invalid Conclusions[a] | Unnecessary Trials | Solutions |
|---|---|---|---|---|
| 1. | Screws | .00 | -.23 | .17 |
| 2. | Suburbia | .02 | -.36 | .34 |
| 3. | Pendulum | .05 | -.12 | .04 |
| 4. | Chemical | .02 | -.24 | .44 |
| 5. | Pencils | -.05 | -.44 | .20 |
| 6. | Cowboys | -.09 | -.22 | .25 |
| 7. | Pegboard | -.19 | -.52 | .63 |

Note. -- Experimental treatment has been partialled out.

[a] Square root of number of invalid conclusions.

tasks (not including the chemical task). The equivalent figures for the pendulum task and the other tasks for unnecessary trials and solutions are .21 and -.03. The picture is the same for the chemical task. The correlations between the chemical task and the other tasks (not including the pendulum task) are .05, .10, and .10 for invalid conclusions, unnecessary trials, and solutions. In contrast, the mean correlations among tasks for each measure, when neither the pendulum task nor the chemical task are included, are .25 for invalid conclusions, .31 for unnecessary trials, and .26 for solutions.

There _is_ a distinct difference between the pendulum task and the chemical task on the one hand and the remaining five tasks on the other. In the former two tasks, the instances are _not_ physically arrayed in front of the subject. He must compose or create the instances as he goes along. The instances are physically arrayed in front of the subject in each of the other five tasks. All the subject must do is select the instance. It was known that whether or not instances are arrayed in front of the subject makes a difference in performance. Bruner, et al., (1956, pp. 92-96) found that subjects did much better on "on-the-board" problems than on "in-the-head" problems. But in advance of this study, it seemed as though the difference might be simply one of difficulty. Apparently, the difference is more far-reaching.

A second conclusion is that number of invalid conclusions is a less reliable and consistent measure than either unnecessary trials or solutions.

The median correlation between two similar problems from the same task on number of invalid conclusions was only .19. The mean correlation among tasks (excepting the pendulum and chemical tasks) on invalid conclusions was .25. Still, Table 9 shows that number of invalid conclusions does tend to be fairly highly related to whether or not problems are solved and, furthermore, the relationships are high for the pendulum task and the chemical task, cases where unnecessary trials shows relatively low correlations with solution.

## Summary

A standardized procedure has been developed for administering problems that involve the attainment of conjunctive concepts from tasks which may be described as small, closed systems. Subjects' instance selection behavior and all of the allowable experimenter behavior is recorded in coded form. The verbal behavior of subjects is recorded verbatum. A series of prompts is included in the presentation procedures to get the subject moving if he stops performing or to get the subject to re-examine an invalid conclusion if he will not do so on his own.

A system for content analysing problem-solving protocols has been developed that stresses molecular categories. The categories are combined into three summary measures; namely, number of invalid conclusions, number of unnecessary trials, and number of solutions. Six factor analyses, as well as a priori considerations, provide a justification for the three summary measures. Interrater reliability coefficients for each of the measures were .90 or better. Internal-consistency reliability for two of the measures -- number of unnecessary trials and number of solutions -- if not the third number of invalid conclusions, were moderate but consistent. Other correlations bearing upon the consistency and stability of the measures were obtained, and, on the whole, were moderately high. It must be noted that all of the data were obtained from a homogeneous sample of 60 first graders. Considerably higher correlations might be expected from older samples, or more heterogeneous samples.

All in all, there is room for confidence that the measures are sufficiently reliable, stable,and consistent that they are capable of reflecting differences among treatments and conditions, at least if relatively large samples are employed.

## 5. DESIGN OF THE EXPERIMENT

This section will set forth the experimental design and the procedures employed in assessing the effect of a training procedure upon problem-solving. The essential part of the design was very simple. One group of children received intensive training with the procedure previously described. Another group received no training or treatment of any kind. When the first group had completed training, both groups were presented with a series of problems, some from tasks that had been employed during training to assess retention, some from tasks none of the children had experience with before to assess transfer. The question is did the children who received training do better?

Sample. The California Test of Mental Maturity, Primary Form, 1957 Edition, was administered to all of the first graders in five of the elementary schools from a suburban community in New Jersey. The test was administered and scored by members of the project staff. Originally the plan had been to sample children from mental age strata. However, experience suggested that the training procedures then being developed and refined were too difficult for low-ability children. For this reason, the sample was restricted to the top third of the mental age distribution.

A total of 60 children from the highest one-third of the CTMM mental age distribtuion were randomly selected and assigned to two groups with the restriction that there be fifteen boys and fifteen girls in each group. As a matter of expediency, there was the further sampling restriction that there be exactly six experimental subjects and six control subjects from each of the five elementary schools. Balancing numbers of subjects by school allowed for equitable division of labor among project staff and made possible a manageable load of problem presentation without confounding influences.

The training group and control group were nearly identical with respect to mental age and chronological age. The experimental group showed a mean CA in months of 79.17 and a CA standard deviation of 3.76. The control group showed a mean CA of 79.03 and a CA standard deviation of 3.86. These figures were based on data obtained in December, at the time the CTMM was given, about five months before the administration of problems. The MA mean in months was 105.22 for the training group and 104.53 for the control group. The respective standard deviations were 9.17 and 8.96.

Training procedure. A subject in the training group met with an experimenter for 20-minute individual training sessions three times per week, unless he was ill or there was a special school program being offered. Training always took place in the morning. The order in which subjects were trained was continually varied according to a predetermined schedule.

Training took place in whatever reasonably private locations the school could provide. These were, namely, a janitor's closet, a coat closet, an art supply room, an alcove to a teacher's room, and a heated garage.

Each member of the project staff -- four research assistants and the principal investigator -- trained six subjects in a certain school. All of the staff had conducted from one to twelve training sessions with each of ten to fifteen children before the training that was part of the experiment began.

As was outlined earlier, training consisted of two phases. In the first phase, problem-solving behavior was shaped on three card array tasks. The second phase involved abbreviated training with five tasks -- a pegboard device, a collection of toy plastic cowboys, a collection of leaf facsimilies, a switching device, and a ball-and paddle device. These five tasks were utilized in counterbalanced orders, with random assignment of subjects to orders.

Procedure for presenting problems. The substantative aspects of procedure were described in the previous section. Physical and administrative arrangements will now be discussed.

Subjects were run one at a time by an experimenter who did not know whether the subject was a member of the training group or the control group.[9] At each session the child performed on both of the problems involving a single task. The length of sessions was variable depending upon the task as well as individual differences. The range was about 10 minutes to 40 minutes and the average about 25 minutes.

---

[9] It was impossible for the experimenter to avoid learning the group identification of some of the subjects as, for instance, when a child volunteered that he had played that game before with Mr.__________.

The cowboy task and the pegboard task were used in training and were presented again approximately one week after the completion of training to assess retention. All subjects performed on the cowboy task and then the pegboard task, in that order.[10] Two problems from five new tasks -- screws, suburbia, pendulum, chemical, and pencils -- were then presented to assess transfer of training. The orders in which these tasks were administered were counterbalanced. Subjects were assigned to orders by school. On any given day, all of the experimental subjects and control subjects in a certain school were given the same task. As will be seen later, there were no significant effects whose interpretation was confounded by this deviation from ideal design. There were three reasons for assigning orders by school instead of by subject. First, the arrangement made the best use of a limited supply of materials and apparatus. Second, the procedure made it possible for each experimenter to concentrate on a single task at a time. Since instructions, procedures, and coding were rather involved, errors may have been avoided as a result. Finally, the procedure made it possible to present a given task to all of the children in a school in a brief period of time, usually a single morning,[11] before there was any opportunity for the children to discuss the task with one another.

---

[10] In retrospect, it would have been a good idea to counterbalance the order of presentation of these two tasks.

Experimenters worked in pairs when presenting problems. On any given day, the two experimenters presented a certain task to the subjects in a single school, each experimenter testing half of the children. After completing testing for the day, the experimenters examined one another's protocols, checking for legibility, proper task and problem identification, proper subject identification, and "legal" use of codes.

Each subject received three tasks from one experimenter and four from another. The order in which children were tested was varied from day to day on a systematic basis, with random assignment of subjects to orders. Three weeks elapsed between the presentation of the first and the last task.

If a child was absent from school on a day when he was to receive a task, the experimenter checked a special list to see if the child was an experimental or a control subject. The experimental subject who was absent was given the task he missed on the first day he returned to school. The order in which he was to receive tasks was maintained. If a control subject was absent, he was replaced with a randomly selected alternate of the same sex for that day only. Correlations were run without alternates. Subsidiary analyses were made without alternates to ascertain whether there were cumulative effects from performing on problems. In a word, the means were the same with or without the alternates.

---

[11] The chemical task took part of the afternoon as well.

Problems were administered in such relatively quiet and private spaces as the schools were able to furnish. All of the spaces employed during training were also used for presentation of problems, along with other, similar spaces.

Statistical analysis. Three-way analysis of variance of a "mixed" (Lindquist, 1953, Type III, pp. 281-284) type was the main technique employed. Treatment, sex, and task were the three factors. Unnecessary trials and invalid conclusions were the dependent variables in the cases where this three-way analysis was used. Number of solutions was not normally distributed by task but was more or less normally distributed when solutions to more than one task were combined. Hence, task was dropped as a factor when number of solutions was the dependent variable. A series of chi-squares was computed to assess the effect of treatment by task on number of solutions.

Subsidiary analyses were performed to assess the effect of order, in a treatment by order design. Several other techniques were employed for minor purposes. These will be described as the need arises.

As has already been indicated, the one per cent has been chosen as the initial value for rejection of the null hypothesis when a parametric test is employed, since there are some departures from normality and, in the case of a couple of the tests to be performed, heterogeneity of variance.

## 6. RESULTS AND DISCUSSION

The main results that will be described and discussed concern the acquisition, retention and the transfer of problem-solving behavior. Other results pertain to the cumulative effects of performing on a series of problems with or without previous training and the effects of training on aptitude factor scores.

Acquisition of problem-solving behavior. Children were allowed a maximum of 27 20-minute training sessions. Time did not permit a greater number of sessions. Table 11 shows the cumulative number of subjects reaching criterion of mastery on the several training tasks within the time available.

For each task, the criterion was the attainment of seven out of eight consecutive concepts in exactly the number of instances required by the ideal method. Neither unnecessary trials nor invalid conclusions were permitted on a problem if it was to be counted toward criterion.

Table 12 presents number of subjects who failed to reach the criterion of mastery by task and by reason for failure. The category "did not attempt for lack of time" means quite literally that 27 training sessions were completed before the child could complete the task. As to "could not master," after a child had spent four sessions on a task, the experimenter could skip on to the next task if, in his judgement, the child was unlikely to master the current task. One thing

## Table 11

Cumulative Number of Subjects Reaching Criterion of Mastery on Training Tasks Within 27 20-Minute Training Sessions

| Reached Criterion on | Cumulative Number of Subjects |
|---|---|
| Second phase training | |
| Five tasks | 16 |
| Four tasks | 26 |
| Three tasks | 28 |
| Two tasks | 30 |
| One task | 30 |
| First phase training | 30 |

Table 12

Number of Subjects who Failed to Reach Criterion of Mastery on Training Tasks

| Tasks | Reason for failure | |
|---|---|---|
| | Did not attempt for lack of time | Attempted but could not master |
| Switching device | 1 | 0 |
| Ball-and-paddle device | 0 | 0 |
| Leaves | 0 | 10 |
| Cowboys | 2 | 1 |
| Pegboard | 3 | 3 |
| Totals | 6 | 14 |

that emerges from Table 12 is that the leaf collection was considerably more difficult than the other tasks. Fully a third of the subjects could not reach criterion on this task. The reason is not hard to find. The leaves involved less salient dimensions and less familiar language than the other tasks. The dimensions were pinate versus palmate, simple versus compound, and notched-edge versus smooth-edge and precisely these words were used with the children.

Table 13 shows median problems to criterion by task and by order. The experimenter was the judge of when the subject was ready to start on the criterion list of problems for a task. Notice that the switching task and the ball-and-paddle task were easier than any of the other tasks. In fact, only four children required as many as one problem to reach criterion on the switching task and only one child required as many as one problem to reach criterion on the ball-and-paddle task. Also, most of the subjects went quickly through the preceding stages of training on these two tasks. Here, it would seem that subjects entered training with virtually the behavior required to solve the problems. Yet high-ability third graders who had not received any training solved only 40 per cent of the switching problems presented to them (Walsh, 1963).

The switching device consisted of a small metal box with a row of six toggle switches across the front. There were two rows of switches on the back of the box with which the experimenter controlled the relevance of the switches on the front. There was also a "master switch"

Table 13

Median Problems to Criterion on Training Tasks

| Problem | Order | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5[b] | All orders |
| Switch | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Ball-and-paddle | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Leaves | NC[a] | 6.5 | NC[a] | 4.5 | 0.5 | 6.0 |
| Cowboy | 5.0 | 8.0 | 2.0 | 6.5 | 6.5 | 5.5 |
| Pegboard | 0.0 | 4.5 | 0.0 | 4.0 | 0.0 | 0.0 |
| All tasks | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

[a] NC = did not reach criterion.

[b] Subjects who could not attempt the fifth task because of lack of time are not included.

(of the push-button type) and a red light on the front. The focus instance always consisted of all switches up. The problem for the child was finding which switches had to be up to make the buzzer sound and the light go on when the master switch was pushed. The ball-and-paddle task was identical in conception to the switching task. Instead of pushing switches, the child pushed paddles into slots. Instead of a light going on and a buzzer sounding, the test was whether a small rubber ball would roll entirely through the box.

At the level of task analysis developed in an earlier section, there is a high degree of similarity between the switching device and ball-and-paddle device on the one hand and the leaves, cowboys, and pegboard on the other. Nonetheless, it now appears that there were important task differences. The switching task and ball-and-paddle required relatively simple instance selection behavior. A subject had only to begin at the left or right and try each switch or paddle in succession. If a switch had to be up or a paddle had to be out, it was returned to its original position before the next trial.

The leaves, cowboys, and pegboard are tasks in which all of the possible instances are physically arrayed in front of the subject. Instance selection is a relatively complex matter with this sort of task. A well-formed instance differs in exactly one respect from a preceding positive instance. In one way or another, the subject must match the values that are not to be changed against a standard -- perhaps a focus instance, perhaps a verbal recipe, perhaps even an image. Instances are really composed instead of selected on the task in which the instances are physically arrayed in front of the subject. Contrast instance selection on the switching task or ball-and-paddle task. As long as the simple rule of returning switches and paddles proved necessary to their original positions is followed, instance selection is merely a matter of manipulating the next switch or paddle. At the conclusion of any trial, the task is always in a "next-instance state," except for the switch or paddle to be manipulated.

Charting progress through a problem is also less difficult on the switching task or the ball-and-paddle task than on the other tasks. A "natural" order in which to test variables is apparent for switches or paddles layed out in a row. The questions what variables have been tried and what variable should be tried next are not so easily answered in tasks in which the instances are physically arrayed in front of the subject.

The hope was to demonstrate the development of learning sets, in the form of decreasing numbers of problems to criterion as the number of tasks increased. As can be seen from Table 13, there was a median of 0.0 problems to criterion on even the first task. It cannot be assumed, as will be seen later, that the bright first grader already knows how to solve problems of this type, so it must be the case that the training was quite successful in achieving its objectives.

Assume for the moment that training on the switching task and ball-and-paddle task was largely irrelevant to performance on the other tasks, that is that there was no transfer from the former to the latter. Table 14 shows that when the switching task and ball-and-paddle task are ignored, the expected decreasing gradient of problems to criterion does appear. Of course, tasks are no longer counterbalanced with respect to order. In an attempt to remove the effect of task, subjects were ranked within each task. Then the two-way analysis of ranks suggested by Friedman (Walker and Lev, 1953, p. 438) was applied to the scores already ranked by task. For what it is worth, an *F*-value of 2.92 ($.10 > P > .05$) was obtained for the effect of order. Recognize that because of the subjective nature of the judgement of when a subject was to start the criterion list of problems for a task, the decreasing numbers of problems to criterion could reflect changes in the experimenter (i.e., he grows progressively more conservative about starting subjects on the criterion list) rather than changes in the subject. On the other hand, one would expect the sharpest decline on the first few tasks, in the present case, the card array tasks that were treated differently and which, therefore, cannot be compared with the other tasks. All and all, whereas the data suggest that on the whole training was quite successful, there are no grounds for conclusions about the development of learning sets.

### Table 14

Median Problems to Criterion on Three Selected Training Tasks by Order of Presentation

| Tasks | Order | | | All orders |
|---|---|---|---|---|
| | 1 | 2 | 3 | |
| Leaves | NC[a] | 6.5 | 2.0 | 6.0 |
| Cowboys | 7.5 | 6.5 | 3.0 | 5.5 |
| Pegboard | 0.0 | 3.5 | 0.0 | 0.0 |
| All tasks | 9.5 | 5.5 | 2.0 | 4.0 |

[a] NC = did not reach criterion.

Effects of sex. Sex differences are frequently found in the age range of the children sampled in the present investigation. Surprisingly enough, no differences appeared between boys and girls on any of the measures of problem solving. Nor did sex enter into any significant interactions. Boys proved to have a slightly higher mean mental age than girls, though the difference was not significant ($t = 1.60$, $df = 58$, $p > .05$). Boys, however, also did slightly better on every measure of problem solving. When the somewhat greater aptitude of the boys is discounted, there remain exactly zero differences between sexes. Since sex was never significant, sex is not included as a factor in any of the tables of means and standard deviations that follow.

Retention of problem-solving behavior. Children who received training unquestionably learned and retained successful problem-solving behavior. Table 15 and Table 16 show the frequency of solutions to cowboy problems and pegboard problems. As can be seen, the training group did significantly better on each task.

Table 17 contains mean unnecessary trials on retention problems by treatment and task. Table 18 presents the analysis of variance of unnecessary trials on retention problems. Treatment and task were the only significant effects. The training group showed fewer unnecessary trials than the control group on both the cowboy problems ($t = 5.26$, $p < .01$) and the pegboard problems ($t = 6.23$, $p < .01$).

Table 15

Number of Solutions to Cowboy Problems

| Number of Solutions | Treatment |  | Both Groups |
|---|---|---|---|
|  | Training Group | Control Group |  |
| Two | 18 | 2 | 20 |
| One | 8 | 7 | 15 |
| None | 4 | 21 | 25 |
| Total | 30 | 30 | 60 |

$X^2 = 24.44, P < .01$

Table 16

Number of Solutions to Pegboard Problems

| Number of Solutions | Treatment |  |  |
|---|---|---|---|
|  | Training Group | Control Group | Both Groups |
| Two | 7 | 1 | 8 |
| One | 11 | 3 | 14 |
| None | 12 | 26 | 38 |
| Total | 30 | 30 | 60 |

$X^2 = 14.20$, $P < .01$

Table 17

Means and Standard Deviations of Numbers of Unnecessary Trials on Retention Problems

| | Treatment | | | | | |
|---|---|---|---|---|---|---|
| | Training Group N = 30 | | Control Group N = 30 | | Both Groups | |
| Tasks | Mean | SD | Mean | SD | Mean | SD |
| Cowboys | 2.60 | 2.61 | 6.43 | 1.38 | 4.52 | 2.81 |
| Pegboard | 8.03 | 4.57 | 12.57 | 1.61 | 10.30 | 4.06 |
| Both tasks | 5.32 | | 9.50 | | 7.41 | |

## Table 18

### Analysis of Variance of Unnecessary Trials on Retention Problems

| Source | df | MS | F |
|---|---|---|---|
| Between subjects | 59 | 19.09 | |
| Treatment (T) | 1 | 525.01 | 52.69* |
| Sex (S) | 1 | 37.41 | 3.75 |
| T X S | 1 | 6.08 | .61 |
| Error (b) | 56 | 9.96 | |
| Within subjects | 60 | 22.34 | |
| Task (P) | 1 | 1003.41 | 168.57* |
| P X T | 1 | 3.68 | .62 |
| P X S | 1 | .08 | .01 |
| P X T X S | 1 | .01 | .00 |
| Error (w) | 56 | 5.95 | |
| Total | 119 | 20.73 | |

*P< .01

The study was not designed in such a way that a very discriminating interpretation of the significant task effect is possible. Indeed, within the constraints imposed by the task analysis, as wide as possible a range of task characteristics was included in order to reflect treatment generality or lack of it. Thus, the tasks differ with respect to materials or apparatus, verbal instructions or "story line," number and type of variables, and number of relevant variables.

Table 19 contains mean invalid conclusions on retention problems by treatment and task. Analysis of variance of invalid conclusions on retention problems appears in Table 20. Treatment, task, and the treatment-task interaction were all significant. The training group gave more invalid conclusions than the control group. This unexpected and undesired consequence needs an explanation. Part of the explanation is, no doubt, that in order for a subject to make a mistake he must "have an idea," that is verbal behavior must be emitted. The belief is that most subjects who received training were engaging in task-related verbal behavior (covert as well as overt) whereas verbal behavior on the part of control subjects occurred less frequently and was less likely to be relevant to solving problems. Recall that statements that were not task-related were not scored.

The significant interaction between task and treatment is not a favorable sign either. On the easier of the two tasks, the cowboy task, the training group produced fewer invalid conclusions than the control

Table 19

Means and Standard Deviations of Numbers of Invalid Conclusions on Retention Problems[a]

| | Treatment | | | | | |
|---|---|---|---|---|---|---|
| | Training Group N = 30 | | Control Group N = 30 | | Both Groups | |
| Tasks | Mean | SD | Mean | SD | Mean | SD |
| Cowboys | .65 | .77 | .70 | .79 | .68 | .77 |
| Pegboard | 1.67 | .90 | .60 | .70 | 1.13 | .95 |
| Both tasks | 1.16 | | .65 | | .90 | |

[a] Square root of number of invalid conclusions.

Table 20

Analysis of Variance of Invalid Conclusions on Retention Problems[a]

| Source | df | MS | F |
|---|---|---|---|
| Between Subjects | 59 | .90 | |
| Treatment (T) | 1 | 7.84 | 9.79* |
| Sex (S) | 1 | .05 | .06 |
| T X S | 1 | .18 | .22 |
| Error (b) | 56 | .80 | |
| Within Subjects | 60 | .72 | |
| Task (P) | 1 | 6.22 | 12.89* |
| P X T | 1 | 9.27 | 19.19* |
| P X S | 1 | .61 | 1.27 |
| P X T X S | 1 | .07 | .14 |
| Error (w) | 56 | .48 | |
| Total | 119 | .81 | |

*$P < .01$

[a] Analysis performed on square root of number of invalid conclusions.

group. On the pegboard task, the more difficult of the two, the training group, produced many more invalid conclusions. Perhaps the interaction can be interpreted as follows: The control subject produced a fairly large amount of task-relevant verbal behavior on the easier of the two tasks, a relatively large proportion of which was inappropriate. Very little task-relevant verbal behavior was elicited by the more difficult task and, as a consequence, the control subject made relatively few inappropriate statements. The experimental subject produced task-relevant verbal behavior on both tasks only, unfortunately, a relatively large proportion was inappropriate on the more difficult task. Another possibility is that during training the subjects learned to guess when in trouble in order to elicit clues from the experimenter. In any event, the data on invalid conclusions are interesting and the issues involved are in need of further investigation.

All things considered, the training group was impressively superior to the control group on the cowboy and pegboard problems, the results with respect to invalid conclusions notwithstanding. The training group solved significantly more problems than the control group, and solved them with significantly greater efficiency.

When viewed in absolute terms, the performance of the training group was less satisfactory. After all, most of the children in the training group had recently reached a fairly stringent criterion of mastery on problems from the cowboy and pegboard tasks. Considering this fact, could the training group have been expected to do better than it did? In an attempt to answer this question, let us discount the subjects who failed to reach training criterion. That some subjects failed to reach criterion is a shortcoming, but one cannot reasonably ask whether behavior was retained when it was not acquired in the first place. Subjects who did not reach the training criterion on the cowboy task did about as well as other subjects. However, five of the six children who did not reach pegboard training criterion failed to solve either of the pegboard problems, while the remaining subject solved one of them. If these subjects are discounted, the frequency of solutions to the pegboard problems for the training group shown in Table 16 is not as far out of line with expectations. Consider also that we are not dealing with immediate retention. Because training tasks were employed in counterbalanced orders, there were intervals of up to seven weeks between the attainment of the training criterion and the presentation of problems to measure retention. When these factors are considered, the performance of the training group on the retention problems was not unsatisfactory, whether viewed relatively with respect to the control group or in absolute terms.

Transfer of problem-solving behavior. Subjects in the training group solved a mean of 5.13 transfer problems. The mean for control subjects was 3.30. There were nine transfer problems altogether, distributed over five tasks. Number of solutions was irregularly distributed by tasks taken singly, but appeared to be normally distributed when solutions to all five tasks were combined. Hence, analysis of variance of solutions to transfer problems was performed to get an overall test of significance. Table 21 shows that the training group solved significantly more transfer problems than the control group. Tables 22 through 26 contain frequency of solution for the training group and the control group on each of the tasks. Differences favoring the training group were significant for suburbia problems and pencil problems, and the difference approached significance on screw problems.

Table 27 contains mean unnecessary trials on transfer problems and Table 28 presents the analysis of variance of unnecessary trials. Treatment, task, and the treatment-task interaction were significant. The training group executed fewer unnecessary trials than the control group As was emphasized before, no particular interpretation of the significant task effect is possible because of the nature of the design and procedures.

Because of the significant treatment-task interaction, $\underline{t}$ tests were computed between training group and control group means for each of the tasks separately. The training group had significantly lower mean unnecessary trials on the screws problems ($\underline{t} = 5.59, P < .01$), the suburbia problems ($\underline{t} = 2.74$, $P < .01$), and the pencil problems ($\underline{t} = 4.03$, $P < .01$). Differences on the pendulum and chemical problems were not significant.

Table 21

Analysis of Variance of Solutions to Transfer Problems

| Source | df | MS | F |
|---|---|---|---|
| Treatment (T) | 1 | 10.08 | 17.33* |
| Sex (S) | 1 | .96 | 1.66 |
| T X S | 1 | .40 | .69 |
| Error | 56 | .58 | |
| Total | 59 | .75 | |

*P < .01

Table 22

Number of Solutions to Screw Problems

| Number of Solutions | Treatment | | Both Groups |
|---|---|---|---|
| | Training Group | Control Group | |
| Two | 23 | 15 | 38 |
| One | 5 | 9 | 14 |
| None | 2 | 6 | 8 |
| Total | 30 | 30 | 60 |

$X^2 = 4.84$, n.s.

Table 23

Number of Solutions to Suburbia Problems

| Number of Solutions | Treatment |  | Both Groups |
|---|---|---|---|
|  | Training Group | Control Group |  |
| Two | 10 | 4 | 14 |
| One | 12 | 7 | 19 |
| None | 8 | 19 | 27 |
| Total | 30 | 30 | 60 |

$X^2 = 8.36,\ P < .05$

Table 24

Number of Solutions to a Pendulum Problem

| Number of Solutions | Treatment | | |
|---|---|---|---|
| | Training Group | Control Group | Both Groups |
| One | 11 | 8 | 19 |
| None | 19 | 22 | 41 |
| Total | 30 | 30 | 60 |

$X^2 = .61$, n.s.

Table 25

Number of Solutions to Chemical Problems

| Number of Solutions | Treatment | | |
|---|---|---|---|
| | Training Group | Control Group | Both Groups |
| Two | 2 | 0 | 2 |
| One | 5 | 5 | 10 |
| None | 23 | 25 | 48 |
| Total | 30 | 30 | 60 |

$X^2 = 2.08$, n.s.

Table 26

Number of Solutions to Pencil Problems

| Number of Solutions | Treatment | | Both Groups |
|---|---|---|---|
| | Training Group | Control Group | |
| Two | 23 | 11 | 34 |
| One | 5 | 10 | 15 |
| None | 2 | 9 | 11 |
| Total | 30 | 30 | 60 |

$X^2 = 10.36$, $P < .01$

### Table 27

### Means and Standard Deviations of Numbers of Unnecessary Trials on Transfer Problems

| | Treatment | | | | | |
|---|---|---|---|---|---|---|
| | Training Group N = 30 | | Control Group N = 30 | | Both Groups | |
| Tasks | Mean | SD | Mean | SD | Mean | SD |
| Screws | 4.73 | 2.85 | 8.20 | 2.16 | 6.47 | 3.03 |
| Suburbia | 7.10 | 2.92 | 8.80 | 2.58 | 7.95 | 2.83 |
| Pendulum[a] | 8.87 | 2.86 | 9.40 | 2.52 | 9.13 | 2.67 |
| Chemical | 14.30 | 2.89 | 13.30 | 2.28 | 13.80 | 2.61 |
| Pencil | 4.27 | 2.61 | 6.77 | 2.34 | 5.52 | 2.74 |
| All tasks | 7.85 | | 9.29 | | 8.57 | |

[a] Twice the score on the pendulum problem was used in order to make this task comparable with the other tasks in which two problems were administered.

Table 28

Analysis of Variance of Unnecessary Trials on Transfer Problems

| Source | df | MS | F |
|---|---|---|---|
| Between Subjects | 59 | 12.02 | |
| Treatment (T) | 1 | 144.21 | 15.01* |
| Sex (S) | 1 | 21.33 | 2.22 |
| T X S | 1 | 5.88 | .61 |
| Error (b) | 56 | 9.61 | |
| Within Subjects | 240 | 18.71 | |
| Task (p) | 4 | 800.56 | 166.46* |
| P X T | 4 | 47.31 | 9.84* |
| P X S | 4 | 3.38 | .70 |
| P X T X S | 4 | 1.94 | .40 |
| Error (w) | 224 | 4.81 | |
| Total | 299 | 17.39 | |

*P < .01

Tables 29 and 30, respectively, contain mean invalid conclusions on transfer problems and analysis of variance of invalid conclusions. Significant Fs were obtained for task and the treatment-task interaction. The training group gave significantly more invalid conclusions ($t = 3.27$, $P < .01$) on the suburbia problems than the control group. Otherwise, none of the task by task differences was significant.

The training group showed generalized transfer of problem-solving behavior. The training group solved significantly more problems new to them than the control group and solved the problems with significantly greater efficiency. This is a finding with important implications, for seldom, if ever, has transfer of training within the area of problem solving been achieved with children.

Though there was no overall difference between the training group and control group in number of invalid conclusions on transfer problems, in one respect the findings are similar to those obtained with the retention problems. The training group showed significantly more invalid conclusions than the control group on the suburbia task. Both suburbia problems involved three relevant variables. Recall that the training group showed much more inappropriate verbal behavior than the control group on the pegboard problems, which also involved three relevant variables. There appears to be a deficiency in the strategy that was learned with respect to problems involving three relevant variables. An explanation of the difficulty was proposed when the pegboard task

Table 29

Means and Standard Deviations of Numbers of Invalid Conclusions on Transfer Problems[a]

| | Treatment | | | | | |
|---|---|---|---|---|---|---|
| | Training Group N = 30 | | Control Group N = 30 | | Both Groups | |
| Tasks | Mean | SD | Mean | SD | Mean | SD |
| Screws | 1.15 | .73 | .89 | .65 | 1.02 | .69 |
| Suburbia | 1.42 | .59 | .90 | .78 | 1.16 | .73 |
| Pendulum[b] | 1.10 | .78 | 1.22 | .53 | 1.17 | .67 |
| Chemical | 1.13 | .59 | 1.43 | .35 | 1.28 | .50 |
| Pencil | .94 | .57 | .85 | .65 | .89 | .60 |
| All tasks | 1.15 | | 1.06 | | 1.10 | |

[a] Square root of number of invalid conclusions.

[b] Twice the (raw) score on the pendulum problem was used in order to make this task comparable with the other tasks in which two problems were administered.

Table 30

Analysis of Variance of Invalid Conclusions on Transfer Problems[a]

| Source | df | MS | F |
|---|---|---|---|
| Between Subjects | 59 | .56 | |
| Treatment (T) | 1 | .83 | 1.46 |
| Sex (S) | 1 | .08 | .14 |
| T X S | 1 | .01 | .02 |
| Error (b) | 56 | .57 | |
| Within Subjects | 240 | .37 | |
| Task (P) | 4 | 1.89 | 5.73* |
| P X T | 4 | 1.42 | 4.31* |
| P X S | 4 | .37 | 1.12 |
| P X T X S | 4 | .20 | .60 |
| Error (w) | 224 | .33 | |
| Total | 299 | .41 | |

*$p < .01$

[a] Analysis performed on square root of number of invalid conclusions.

was discussed. The matter is in need of further investigation.

Note that the training group did distinctly better than the control group on the screws, pegboard, and pencils, but not on either the pendulum problem or the chemical problems. What was different about the later two tasks? One answer is that the instances were not physically arrayed in front of the subject. The subject had literally to create or construct the instances.[12] The job of constructing instances required a repertoire of behavior that was not learned during training. Apparently, children learned a "system processing strategy" that was evoked by a limited class of tasks. If it were also possible to teach "system defining" behavior, then the child would possess a repertoire of skills of extraordinary social utility.

How shall "system defining" operations be conceived. Evidently, they deal with such issues as what are the variables, what are the states of each variable, what are the unique cells generated in a matrix of variables, and what are the relationships that could obtain among variables. Piaget's notions of "class inclusion operations" and "qualitative class multiplication" are unquestionably relevant (e.g., Inhelder and Piaget, 1958; Flavell, 1963).

---

[12] Instances were not physically arrayed in front of the subject in the switching task or ball-and-paddle task either. But in these tasks, the system was in a "next-instance state" so long as some very simple rules were followed. Whereas the pendulum task and chemical task required creating the next instance in a very real sense.

*Effects of order.* Since the design of this study involved a no-treatment control group, it is important to evaluate whether or not there was facilitation from merely performing on a series of problems. Tables 31, 32, and 33 contain analyses of the effect of order on the three dependent variables. Treatment was included so as to see if there were significant interactions. Analysis of variance was applied to solutions because when categorized by order, the forms of the distribution are at least homogeneous if not normal. As can be seen, order was not significant as a main effect nor did order enter into any significant interaction. When the error term used to test order effects is reduced as though task had been included in the analysis, order is still not significant. Since order was insignificant, tables of means and standard deviations were omitted. Evidently merely performing on a series of problems does *not* have a facilitating effect.

*Effect of training on aptitude factor scores.* The California Test of Mental Maturity was given to 400 first graders, including the 60 children who took part in the experiment, before training began. Subtest scores were intercorrelated. The correlation matrix was factor analysed by the principal components method. Two different estimates of communality were tried. The first estimate used was the squared multiple correlation. The analysis based on this estimate proved unstable, the first factor accounting for more than 100 per cent of the common variance. The value ultimately used as the estimate of communality was

Table 31

Treatment by Order Analysis of Unnecessary Trials on Transfer Problems

| Source | df | MS | F |
|---|---|---|---|
| Between Subjects | 59 | 12.02 | |
| Treatment (T) | 1 | 144.21 | 14.80* |
| Error (b) | 58 | 9.74 | |
| Within Subjects | 240 | 18.71 | |
| Order (O) | 4 | 1.91 | .11 |
| O X T | 4 | .87 | .05 |
| Error (w) | 232 | 19.31 | |
| Total | 299 | 17.39 | |

*$p < .01$

## Table 32

### Treatment by Order Analysis of Invalid Conclusions on Transfer Problems[a]

| Source | df | MS | F |
|---|---|---|---|
| Between Subjects | 59 | .56 | |
| Treatment (T) | 1 | .83 | 1.51 |
| Error (b) | 58 | .55 | |
| Within Subjects | 240 | .37 | |
| Order (O) | 4 | .84 | 2.08 |
| O X T | 4 | .22 | .55 |
| Error (w) | 232 | .37 | |
| Total | 299 | .41 | |

[a] Analysis performed on square root of number of invalid conclusions.

Table 33

Treatment by Order Analysis of Solutions on Transfer Problems

| Source | df | MS | F |
|---|---|---|---|
| Between Subjects | 59 | .75 | |
| Treatment (T) | 1 | 10.08 | 17.22* |
| Error (b) | 58 | .59 | |
| Within Subjects | 240 | .72 | |
| Order (O) | 4 | .38 | .54 |
| O X T | 4 | .02 | .02 |
| Error (w) | 232 | .73 | |
| Total | 299 | .72 | |

*$p < .01$

the highest correlation. Of the factors extracted, the first four were preserved for rotation by the Biquartimin method (Harmon, 1960). Together these four factors accounted for 96 per cent of the common variance and the total exceeded 100 per cent when a fifth factor was included. The first three factors were interpreted, namely, as General Reasoning, Numerical Facility, and Verbal Comprehension (Fraunfelker and Anderson, 1964). The CTMM was administered again after the experiment to all subjects. Factor scores (composites) were computed for both the pre-experimental and post-experimental administration on the basis of the factor analysis outlined.

The question is was there transfer of training to the aptitude factor scores? In principle, it should be possible to improve aptitude by deliberate training. But it must be emphasized that the test was made with the hope rather than the expectation of obtaining a positive result. Multivariate analysis of covariance (Cooley and Lohnes, 1962) was used to test the significance of the difference between the training group and the control group in posttest aptitude factor scores. Covariance adjustments were made for the four pretest aptitude factor scores and chronological age. A lambda of .95 and an $\underline{F}$ of .69 were obtained. Since the result was not significant, none of the other statistics will be presented.

# 7. SUMMARY AND CONCLUSIONS

The chief goal of the project described in this report was to develop a procedure for teaching children the Method of Difference, the problem-solving strategy in which each factor is varied in succession while all other factors are held constant. What exactly is involved in the Method of Difference? What are the components of the method and in what order must the component operations be performed? What is the scope of the method and what are its limitations? Answers to these questions comprise a prescription that can furnish guidelines in the construction of a teaching program.

The Method of Difference is applicable to any task that can be defined as a closed system within which the problem is to attain a conjunctive concept. It is the most efficient method there is for attaining conjunctive concepts within systems that involve relatively few variables. It is most efficient in the sense that the maximum number of empirical tests or trials that will be required to attain a concept is minimized. The Method of Difference requires as many trials as there are variables. Any method that involves manipulating more than one variable at a time will, on the average, require more trials than there are variables, assuming a random sequence of concepts to be attained.

Terminal Method of Difference behavior consists of a sequence of exactly $\underline{k} + 1$ overt responses for each problem where $\underline{k}$ is the number of variables in the task from which the problem is selected. There are to be *k* choices of instances and a conclusion such that the instances that were chosen imply the conclusion that is stated and no other.

Interestingly enough, there is an indefinite number of strategies that could yield terminal Method of Difference behavior. Conclusions may be based on positive instances only, negative instances only, or both positive and negative instances. Conclusion drawing may begin after the selection of the first instance or not until all of the instances have been selected. Conclusions may deal with common elements, necessary values, irrelevant variables, possible concepts, or combinations of these. Instance selection may follow from conclusion drawing, or there may be a relatively autonomous instance selection routine. Progress through the problem may be charted from other functions or there may be a separate index function that can take one of several different forms. In view of the many strategies that could yield the same ultimate, observable behavior, there would appear to be distinct limits upon what can be learned about problem solving from naturalistic study and the process tracing experiment. One good way -- perhaps the only satisfactory way -- that strategies such as those conceived by Hovland (1952) and Huttenlocher (1962) can be differentiated is by teaching them and then analysing the resulting performances.

The job of constructing a program to teach a problem-solving strategy requires analysis at a molecular as well as a molar, prescriptive level.

At a molecular level, a strategy can be conceived as a behavioral chain. A sequence of behavior is emitted under the control of task and problem cues and also under the control of response-produced cues.

A teaching program embodying these conceptions was developed that consisted of seven large units or "games." Each game was designed to teach one or more components of the Method of Difference. The first three games taught chiefly appropriate conclusion drawing behavior. The last four games taught instance selection behavior and coordinated conclusion drawing with instance selection.

There were two phases of instruction. The first phase involved developing Method of Difference behavior in a very deliberate fashion with three card array tasks. The second phase consisted of an abbreviated form of the program with each of five additional tasks. The purpose of the second phase of instruction was to vanish the control of particular task and problem characteristics and bring Method of Difference behavior under the control of abstract, schematic attributes of tasks and problems.

## The Experiment

Two groups of thirty children were randomly selected from among the first graders in five large elementary schools in a suburban community in New Jersey. Each group contained fifteen boys and fifteen girls. The sample was restricted to children in the upper one-third of the mental age distribution, since experience suggested that the program might be too difficult for slow and average children. Mental age data were obtained from the California Test of Mental Maturity (CTMM), Primary Form, 1957 Edition, which was administered and scored by members of the project staff.

One group received the Method of Difference program. The other group got no treatment. Unlike most programed instruction, the teaching was done by humans instead of books or machines.

Following instruction both the training group and the control group took the CTMM again and performed on a series of problems from seven tasks. Two of the tasks were employed during training and, thus, provided measures of retention. Five of the tasks were new to all children. These were intended to assess transfer of training. The transfer tasks were presented in counterbalanced orders.

Experimenters were not acquainted with the children to whom they administered problems. Nor did the experimenters know which of the children were members of the training group and which were members of the control group.

Problems were administered in a standardized manner. Subjects were first oriented to the task, meaning that they acquired or demonstrated knowledge of a list of concepts that included the problem concept. Each problem involved the attainment of a conjunctive concept. The experimenter presented a "focus instance," that is, an exemplar of the concept. The child then selected or constructed instances until he could name the concept. Each time the child chose an instance, he learned immediately whether the instance showed the concept (positive instance) or did not show the concept (negative instance). On some tasks, the feedback was the physical consequence of the child's action. On others, the experimenter gave feedback orally. If a child stopped trying to solve the problem or stated an incorrect conclusion with which he was satisfied, the experimenter attempted to keep him going with one of a series of standardized prompts. The experimenter was supposed to keep the child performing if he could until either the child solved the problem or selected the maximum number of instances allowed on the problem. In order to solve a problem, a child had to select a series of instances and state a conclusion such that the instances implied the conclusion and no other.

The experimenter wrote a protocol recording the instances the child selected, the conclusions the child stated, and the statements the experimenter made in the sequence in which they occurred. The instances and permissable experimenter statements were recorded in coded form, resulting in a savings in experimenter time such that he could concentrate on proper administration of the problem. Immediately following a day's administration of problems, the two experimenters working in the same elementary school exchanged protocols to check them for legibility, completeness, accuracy of subject and problem identification, and "legal" use of codes.

The area of problem solving has been plagued with inadequate measurement. Noting success or failure on one or two long, complex problems is likely to be unreliable and involve floor or ceiling effects, particularly when training is presented. Such measures as time to solution and number of hints required have tangential validity. Assessments of complex patterns of problem solving typically consist of more or less subjective rating or classification.

The procedure employed in the present study "externalized" much of the important problem-solving behavior, making possible improved measurement. The experimenter has control of the sequence of instances in the classical concept formation study. Subjects selected or created instances in the present study, so the sequence of instances, as well as the number, is a datum.

A "trial" is defined as the selection of an instance. A trial is inefficient if it yields less than the maximum attainable amount of information. A trial is redundant if it yields a conclusion implied by previous trials. Finally, a trial is repetitious if it exactly duplicates a previous trial. The sum of inefficient, redundant, and repetitious trials constitutes the first measure of problem solving, number of unnecessary trials. Unqualified declarative statements about the concept or the relevance of variables can be classified as inconsistent, consistent, bizarre, and valid. The sum of inconsistent, consistent and bizarre statements about the concept comprised number of invalid conclusions, the second measure of problem solving. If a subject selected a set of instances and stated a conclusion such that the instances implied the conclusion and no other, then he solved the problem. Number of solutions was the third measure of problem solving

Scores on the three summary measures of problem solving were obtained in a content analysis of problem-solving protocols. Protocols, identified only by number, were scored by two raters working independently. Interrater reliability coefficients averaged .97, .90, and .91 per problem for unnecessary trials, invalid conclusions, and solutions, respectively. Two similar problems were presented for each of the seven tasks, except the pendulum task. To assess reliability, correlations between scores on the two problems were obtained separately for the training group and the control group, and then averaged so as to avoid a spurious treatment effect. The average reliabilities by task, estimated with the Spearman-Brown formula, were .50, .32, and .53 for unnecessary trials, invalid conclusions, and solutions, respectively. While not of the magnitude that might be hoped, the reliabilities exhibited by number of unnecessary trials and number of solutions, if not number of invalid conclusions, are sufficiently high that differential treatment and condition effects can be obtained, at least when several tasks and relatively large samples are employed. Recognize that these coefficients are based on a very homogeneous sample of first graders. Higher coefficients probably would be obtained from more heterogeneous samples or older samples.

## Results

All thirty of the children who received training reached a criterion of mastery on the first phase of training, which involved three card array tasks. Only sixteen subjects reached criterion on all five of the tasks comprising the second phase of training; however, twenty-six of them reached criterion on four tasks. Just one task gave subjects appreciable difficulty; namely, the collection of leaf facsimiles. This task was undoubtedly difficult because it involved subtle variables and unfamiliar names.

The hope had been to demonstrate cumulative transfer of training from task to task. Unfortunately, two of the tasks -- the switching device and the ball-and-paddle device -- proved so simple that almost all of the children performed at the criterion level to begin with. These two tasks seemed to be easy because a simplified instance selection algorithm could be employed. A suggestion of cumulative transfer appeared on the remaining three tasks, but the trend was not significant -- nor would the procedures have permitted a confident inference in any event.

The performance of the training group was distinctly superior to the performance of the control group on problems presented to measure retention. The training group solved more problems and solved these problems with fewer unnecessary trials. At first glance, it appears that the training group did not do as well in absolute terms as might be hoped. But when those subjects who did not reach the training criterion are discounted and when the fact that the problems were administered as long as seven weeks after the time that the training criterion was reached is considered, then the performance of the training group, either relative to the control group or in absolute terms, reaches expectations.

The most striking result of the study was the attainment of positive, nonspecific transfer of training. The training group solved many more problems from tasks entirely new to them -- involving different variables, different materials and apparatus, and, to some extent, different language -- than the control group. Moreover, the training group solved problems more efficiently than the control group, exhibiting fewer unnecessary trials.

The training group failed to do any better than the control group on the pendulum problem and the chemical problems. The instances were not physically arrayed in front of the child when he performed on these problems.

The training group showed more invalid conclusions on the pegboard task (retention) and the suburbia task (transfer); otherwise, there were no differences between groups in terms of invalid conclusions. Problems from each of these two tasks involved four variables, three of which were relevant. It may be that the training group showed more invalid conclusions because on all problems it emitted more task-relevant verbal behavior, and, when the problem was difficult, a large per cent of the behavior was inappropriate. Or perhaps there was a deficiency in the training, such that children learned to guess when in trouble, so as to elicit clues from the experimenter.

Contrary to the usual pattern at this age level, there were no differences between boys and girls on any of the measures of problem solving, nor did sex interact with either treatment or task.

There were no differences in performance on any of the measures as a function of order of presentation, nor did order interact with treatment. The failure to find differences attributable to order indicates that merely performing on a series of problems is not facilitating, and suggests that the effects of training were neither transitory nor superficial.

Finally, there were no differences between the training group and the control group on a set of four aptitude factor scores, based on an intelligence test administered after the training was completed, when adjustments were made for age and pretest aptitude factor scores.

## Implications

The results of the study show that children can acquire, retain, and transfer rather complex problem-solving behavior. This finding is noteworthy, especially with respect to transfer, since as a general rule, previous research suggests that at best children will achieve a rather tenuous grasp of behavior that seems to involve complex, abstract operations. Ausubel and Schiff (1954) found that young children taught to predict the operation of a teeter-totter on the basis of the principle that the longer side from the fulcrum falls (when both sides are equally weighted), later showed little resistance to the reasoning that the determining factor was the color of the blocks used as weights. Older children resisted the spurious reasoning. Smedslund (1961a) showed that children could be taught the principles of conservation of weight and substance, that is, that changing the shape of an object does not alter its weight or volume. He went on to show (1961b) that children who learned the concept by "external reinforcement" (e.g., they were taught) reverted to non-conservation explanations when by slight-of-hand they were presented with evidence that contradicted the principle of conservation. About half of the children who already knew the principle when training began resisted interference. Ervin (1960) in a study that has already been described, taught third and fourth graders a procedure for hitting a target on the rebound by aiming at a backstop, but failed to obtain transfer to a task that required aiming a light at a mirror in such a way that the light would reflect on a target.

Though some of the persons leading the curriculum reform movement, such as Page (quoted in Bruner, 1960, p. 126), apparently believe that young children are more capable intellectually than adults, the weight of evidence is unquestionably to the contrary. Intellectual operations develop between infancy and adolescence that have important effects on learning, retention, and transfer of training. Piaget's interpretation of cognitive development in terms of four stages is pre-eminent. There are many who have reservations about Piaget's work, particularly with regard to methodology. In the main, nonetheless, his outline of development has been substantiated by other researchers (e.g., Peel, 1959), even by those employing different and more rigorous techniques (Braine, 1959).

What are the implications for instruction of the evidence on intellectual development? The prevailing view is that instruction, both content and method, should be matched to the child's level of intellectual functioning. From approximately six to twelve years of age, children are at what Piaget calls the stage of concrete operations. Children at this stage are poorly equipped to comprehend formal, verbal propositions. Learning that involves abstract propositions -- directly or as intervening, mediating steps -- is likely to be unstable and tied to the specific circumstances of instruction. The studies described in the preceding paragraph are often cited to show the frailties of learning outcomes when instruction is attempted that is too ambitious considering the child's stage of intellectual development.

Without doubting the general validity of the developmental evidence, a different interpretation may be placed upon the failure of children to acquire easily complex operations. Those skills that comprise the "adult level" of intellectual functioning can be taught to children so that they too are characterized by swift learning and spontaneous transfer of training.[13] Since it appears to take about fifteen years for these skills to develop when learning is left to incidental contingencies, it is to be expected that development would take quite a while and involve many steps even under optimum conditions of instruction. Thus, it is not surprising that the "advanced" behavior children acquire in short-term studies such as those of Smedslund (1961), Ervin (1960), or Ausubel and Schiff (1954), is neither permanent automatically generalized, or impervious to interference.

The Method of Difference program does involve concrete materials and devices. In this respect, it is consistent with the stage of intellectual development of the six- and seven-year old children who received training. But with respect to the concepts taught, the Method of Difference program should be too advanced for six- and seven-year olds. Inhelder and Piaget (1958, p. 335) have stated that "the adolescent superimposes propositional logic on the logic of classes and relations.

---

[13] The assumption is that maturation in the physical sense is less critical than Gesell or even Piaget has argued. For the most part, contemporary writers are inclined to discount maturation as an important factor in cognitive development after the first few years of life. Ausubel (1963, p. 114), for instance, has asserted that "it is erroneous to believe that stages of intellectual development are exclusively the products of 'internal ripening,' and hence that they primarily reflect the influence of endogeneous factors. Gesell's embryological model has little applicability to human development beyond the first year of life when environmental factors become increasingly more important ... "

Thus, he gradually structures a formal mechanism (reaching an equilibriam point at about 14-15 years) which is based on both the lattice structure and the group of four transformations. This new integration allows him to bring inversion and reciprocity together into a single whole. As a result, he comes to control "... hypothetico-deductive reasoning and experimental proof based on the variation of a single factor with the others held constant (all other things being equal) ..." There is a temptation to conclude that the project achieved what Piaget says is impossible. Six- and seven-year old children learned the strategy of varying each factor in succession while holding all others constant. Mastery of this strategy should not be possible until fourteen or fifteen.

A more circumspect conclusion is actually indicated. The problem-solving performance of children who received training improved, but the nature of the strategies that the children acquired cannot be specified with certainty. At several crucial points, instruction could have permitted alternative strategies to develop. The behavior that evolved at these junctures was a function of entering behavior or more or less accidential contingencies. Based on an analysis of the program and observation of children during training, the belief is that most of the children acquired what may be called a "common elements strategy." With such a strategy the conclusion consists of those attributes that the positive instances share. There are several versions of the common elements strategy (e.g., Figure 6) that yield terminal Method of Difference behavior.

Clearly, Inhelder and Piaget had in mind a Method of Difference strategy in which conclusions are based on necessary values and irrelevant variables. Thus, there seems to be a difference between what Inhelder and Piaget believe is involved when adolescents or adults exhibit Method of Difference behavior and the strategy that was probably acquired by most of the children who received training in the present project.

It should be possible to teach children a Method of Difference strategy like the one Inhelder and Piaget envision. Indeed, a program designed for just this purpose has already been developed. Though the program has not been tried yet with children, based on experience it looks as though it has a good chance of success.

Recall that children who received training failed to do any better than control children on tasks in which the instances were not physically arrayed in front of the subject. In the language of Section 2, the children proved incapable of "defining the systems" that these tasks embodied. On the basis of the developmental literature, one would not expect primary grade children to have the repertoire of skills involved in conceiving the interaction of variables. Hunt (1960, pp. 231-232) has illustrated nicely the point as follows: "Faced with the problem of describing the population of animals on a newly discovered planet, the child with only concrete operations could merely do the empirical task of searching for animals and assigning them to ... classes ... On the other hand, if an adolescent with formal operations were faced with such a task, he would be capable of considering all of the various classes of animals that are conceivable, and he might set up a table of these possibilities before he began his exploration."

We may be reasonably certain, on the basis of the developmental literature, that six- and seven-year old children are not good at conceiving the classes that may be formed by combining variables. It should be possible to analyse the behavior that is involved in "conceiving combinatorial possibilities" and design instruction to arrange the necessary behavior. If it were possible to teach children Method of Difference behavior and the behavior required to conceive incompletely observable systems, then the children would possess skills of considerable utility for coping with a range of "real life" problems, such as those encountered in school curricula, particularly those new curricula emerging in mathematics and the sciences.

Many of the prominent figures in child development believe that basic changes in level of intellectual functioning are more matters of "internal equilibration" than of learning as the result of "external reinforcement" (e.g., Smedslund, 1961b; Bruner, 1961). The present project was based on the distinctly different presumption that reasoning and problem-solving capabilities consist of repertorires of behavior, modifiable in accordance with principles of learning, instead of inaccessible structures and processes. The results of the project suggest that instruction based on this assumption is feasible.

# REFERENCES

Anderson, R. C. Manual for presenting logical problems. New Brunswick, N. J.: Cognition Project, Rutgers -- The State University, 1963.

Anderson, R. C., & Anderson, R. M. Transfer of originality training. J. educ. Psychol., 1963, 54, 300-304.

Ausubel, D. P. The psychology of meaningful verbal learning. New York: Grune & Stratton, 1963.

Ausubel, D. P., & Schiff, H. M. The effect of incidental and experimentally induced experience in the learning of relevant and irrelevant causal relationships by children. J. genet. Psychol., 1954, 84, 109-123.

Braine, M. D. S. The ontogeny of certain logical operations: Piaget's formulations examined by nonverbal methods. Psychol. Monogr., 1959, 73, No. 4 (Whole No. 475).

Bruner, J. S. The process of education. Cambridge, Mass.: Harvard University Press, 1960.

Bruner, J. S. The act of discovery. Harv. educ. Rev., 1961, 31, 21-32.

Bruner, J., Goodnow, Jacqueline, & Austin, G. A study of thinking. New York: Wiley, 1956.

Callantine, Mary F., & Warren, J. M. Learning sets in human concept formation. Psychol. Rep., 1955, 1, 363-367.

Conant, James B. Modern science and modern man. New York: Columbia University Press, 1952.

Cooley, W. W., & Lohnes, P. R. Multivariate procedures for the behavioral sciences. New York: Wiley, 1962.

Crafts, L. W. Routine and varying practice as preparation for adjustment to new situations. Arch. Psychol., N. Y., 1927, 14, No. 91.

Dashiel, J. F. An experimental isolation of higher level habits. J. exp. Psychol., 1924, 7, 391-397.

Duncan, C. P. Transfer after training with single versus multiple tasks. J. exp. Psychol., 1958, 55, 63-72.

Duncan, C. P. Effects of instructions and information on problem solving. J. exp. Psychol., 1963, 65, 321-327.

Eckstrand, G. A., & Wickens, D. D. Transfer of perceptual set. J. exp. Psychol., 1954, 47, 274-278.

Ervin, Susan M. Training and a logical operation by children. Child Develpm., 1960, 31, 555-563.

Flavell, J. The developmental psychology of Jean Piaget. Princeton, N. J.: Van Nostrand, 1963.

Fraunfelker, G. H., & Anderson, R. C. An appraisal of the California Test of Mental Maturity for Primary Grade Children. Urbana: University of Illinois, 1964.

Gagné, R. M., & Brown, L. T. Some factors in the programing of conceptual learning. J. exp. Psychol., 1961, 62, 313-321.

Harlow, H. F. The formation of learning sets. Psychol. Rev., 1949, 56, 51-65.

Harmon, H. H. Modern factor analysis. Chicago: University of Chicago Press, 1960.

Hovland, C. I. A "communication" analysis of concept learning. Psychol. Rev., 1952, 59, 461-472.

Hovland, C. I. Computer simulation of thinking. Amer. Psychologist, 1960, 15, 687-693.

Hunt, J. Intelligence and experience. New York: Ronald, 1961.

Huttenlocher, Janellen. Some effects of negative instances on the formation of simple concepts. Psychol. Rep., 1962, 11, 35-42.

Inhelder, Bärbell, & Piaget, J. The growth of logical thinking. New York: Basic Books, 1958.

Kurtz, K. H. Discrimination of complex stimuli: the relationship of training and test stimuli in transfer of discrimination. J. exp. Psychol., 1955, 50, 283-292.

Lindquist, E. F. Design and analysis of experiments. Boston: Houghton Mifflin, 1953.

Meyers, R. E. An experimental program in creative thinking. Minneapolis: Bureau of Educational Research, University of Minnesota, 1961.

Morrisett, L., & Hovland, C. I. A comparison of three varieties of training in human problem solving. J. exp. Psychol., 1959, 58, 52-55.

Newall, A., Shaw, J. C., & Simon, H. A. Elements of a theory of human problem-solving. Psychol. Rev., 1958, 65, 151-166.

Peel, E. A. Experimental examination of some of Piaget's schemata concerning children's perception and thinking and a discussion of their educational significance. Brit. J. educ. Psychol., 1959, 29, 89-103.

Piaget, J. Logic and Psychology. New York: Basic Books, 1957.

Quine, W. V. O. Word and object. Cambridge: Massachusetts Institute of Technology Press, 1960.

Skinner, B. F. The science of learning and the art of teaching. Harv. educ. Rev., 1954, 24, 86-97.

Skinner, B. F. Verbal behavior. New York: Appleton-Century-Crofts, 1957.

Smedslund, J. The acquisition of conservation of substance and weight in children. II. Scand. J. Psychol., 1961, 2, 71-84. (a)

Smedslund, J. The acquisition of conservation of substance and weight in children. III. Scand. J. Psychol., 1961, 2, 85-87. (b)

Snygg, D. Another look at learning theory. Educational Psychologist, 1963, 1, No. 1.

Staats, A. W. Verbal and instrumental response hierarchies and their relationship to problem-solving. Amer. J. Psychol., 1957, 70, 442-446.

Stebbing, M. A. Analysis of Mr. Mill's System of Logic. London: Longmans, Green, 1888.

Thustone, L. L. Psychological implications of factor analysis. *Amer. Psychologist*, 1948, *3*, 402-408.

Thurstone, Thelma. *Learning to think series*. Chicago: Science Research Associates, 1949.

Wolfle, D. L. The relative efficiency of constant and varied stimulation during learning. III. The objective extent of stimulus variation. *J. comp. Psychol*., 1936, *22*, 375-381.

Walker, Helen, & Lev, J. *Statistical Inference*. New York: Holt, 1953.

Walsh, Nancy E. *The relationship between performance on certain long problems and intelligence factors*. Unpublished doctor's thesis. Rutgers -- The State University, 1963.